J Clin Epidemiol Vol. 46, No. 11, pp. 1243–1248, 1993
Printed in Great Britain

0895-4356/93 $6.00 + 0.00
Pergamon Press Ltd

KNOWLEDGE OF MEDICAL HISTORY INFORMATION AMONG PROXY RESPONDENTS FOR DECEASED STUDY SUBJECTS

ALLISON TEPPER,[1]* L. BARBARA CONNALLY,[1] PATRICIA HALTMEIER,[2] EMILY SMITH[2] and MARIE H. SWEENEY[1]

[1]Division of Surveillance, Hazard Evaluations, and Field Studies, National Institute for Occupational Safety and Health, Cincinnati, OH 45226 and [2]Occupational Health Service, New Jersey Department of Health, Trenton, NJ 08625, U.S.A.

(Received in revised form 9 June 1993)

Abstract—Proxy respondents were interviewed for 96 decedents in an occupational cohort. A second respondent was interviewed for 59 decedents. Medical records were reviewed to validate questionnaire information. The percentage of respondents who answered "don't know" (non-response) to questions about medical condition ranged from 5% (cancer and heart disease) to 17% (ulcers). Non-response rates were lowest among spouses, intermediate among children, parents, and siblings, and highest among other relatives and friends. Among 41–55 pairs, depending on the condition, agreement between paired respondents was excellent ($\kappa > 0.75$) for ulcers, cancer, diabetes, and lung disease. A higher percentage of medical records was obtained for decedents with spouse respondents and for decedents with more recent dates of death. Sixty percent or more of the medical records were obtained for patients with cancer ($n = 30$), heart disease ($n = 26$), stroke ($n = 9$), and liver disease ($n = 10$). The positive predictive value of the proxy respondent information for these conditions was 93, 81, 78, and 60%, respectively.

Epidemiologic methods Medical records Proxy respondents Agreement

INTRODUCTION

Proxy respondents are often used in epidemiologic studies involving decedent cases with rapidly fatal diseases and elderly subjects incapable of providing information about themselves. There are several reports on the quality of proxy respondent information about demographic characteristics, personal habits, and occupational history [1–10]. Comparatively fewer reports on the quality of medical history information are available [4, 11–14]. The extant studies examined proxy respondent information for living index cases only [4, 11, 12] or were limited to evaluating proxy respondent information about cancer [13, 14]. The present study of proxy respondents for deceased members of a worker cohort provides new information about the use of proxy respondents to obtain information about a variety of medical conditions among deceased study subjects. Our primary objective was to determine whether different types of proxy respondents can provide accurate information about medical conditions.

*All correspondence should be addressed to: Allison Tepper, Ph.D., Division of Surveillance, Hazard Evaluations, and Field Studies, NIOSH, R-10, 4676 Columbia Parkway, Cincinnati, OH 45226, U.S.A.

METHODS

Study population

A cohort of 586 workers with at least 1 day of employment after 1955 in one of two similar chemical manufacturing facilities in New Jersey

and Missouri was identified for study [15]. In a cross-sectional study of living cohort members, a wide range of health effects potentially related to occupational exposures was examined. Proxy respondents for the 147 workers known to be deceased at the start of the cross-sectional study form the population for the analysis reported here. (Determination of vital status was complete for 93% of the cohort.) Ninety-seven percent of the decedents were male. Their median age at death was 56 years.

The potential first proxy respondent was identified from the death certificate. When the first respondent was deceased or lost to follow-up, funeral directors were contacted and newspaper obituaries were searched for names of other family members or friends. When more than one potential first respondent was identified, the priority order for contact was spouse, offspring and siblings, other relatives, and friends. Motor vehicle license records in New Jersey and Missouri, postal records, and city directories were searched for current addresses.

The identity of a potential second proxy respondent was requested during the first respondent interview. The interviewer explained that it was important to collect information from two people and asked for the name of another person who knew the decedent well.

Data collection

Proxy respondent data were collected from January 1989 through December 1989. Letters were sent to potential respondents to explain the study and alert them to our pending 'phone calls. When subjects could not be contacted by 'phone after the first mailing, two subsequent mailings were conducted, each separated by at least eight attempts to establish contact by 'phone. The third letter was sent by certified mail.

One of two trained interviewers administered a standardized questionnaire to study participants. The questionnaire was administered by telephone. All respondents spoke English. Interview topics included the decedent's demographic characteristics, employment history, personal habits, and medical history. The medical history included heart disease, high blood pressure, liver disease, lung disease, stroke, cancer, diabetes, and ulcers (Table 1).

The interviewers also requested medical records from physicians, clinics, and hospitals for all occurrences of the eight conditions reported by respondents. Medical records were also requested for the last known hospitalization for any conditions identified on the death certificate. A letter sent to medical care providers requested the complete medical record for the decedent whose name, dates of birth and death, social security number, and approximate dates of medical tretament were provided. Follow-up 'phone calls were made when records were not received to determine availability.

Trained abstractors reviewed medical records to confirm the diagnosis reported by the respondents. Abstractors recorded occurrences of any of the conditions of interest mentioned in the medical record.

Data analysis

The ability of respondents to provide information was evaluated by computing rates of non-response to questions about each medical condition. An answer of "don't know" to any question was considered a non-response. For decedents with two proxy respondents, agreement between respondents was measured by the kappa statistic [16]. Kappa values greater than 0.75 are considered excellent agreement, 0.40–0.75 good agreement, and less than 0.40 poor agreement [16]. For each medical condition, analysis of agreement included only pairs for which both respondents answered "yes" or "no".

The extent to which medical records supported the positive report of a medical condition was assessed using only information obtained from the first respondent. The positive predictive value (PPV) was defined as the proportion of confirmed cases among those for whom medical records were retrieved. Decedents with more than one condition were counted as a case for each reported condition.

Table 1. Questions used to obtain medical history information from proxy respondents for deceased study subjects*

1. Was (deceased) ever told by a doctor that he had (condition)?
2. What kind of (condition) did the doctor say it was?†
3. In what year did a doctor first diagnose (condition)?
4. Can you tell me the name and address of the doctor who diagnosed (condition)?
5. Was he ever hospitalized overnight for (condition)?
6. Can you tell me when and where he was hospitalized?

*The medical history covered heart disease, high blood pressure, liver disease, lung disease, stroke, cancer, diabetes, and ulcers.

†This question was asked only for heart disease, liver disease, lung disease, and cancer.

Table 2. Non-response rates for medical condition questions by type of proxy respondent*

	Spouse ($n = 50$)		Child/ Parent ($n = 61$)		Sibling ($n = 31$)		Other ($n = 13$)		Total ($n = 155$)	
	No.	%	No.	%	No.	%	No.	%	No.	%
Cancer	0	0	4	7	1	3	2	15	7	5
Heart disease	0	0	8	13	1	3	2	15	7	5
Diabetes	0	0	3	5	3	10	6	46	12	8
Stroke	0	0	8	13	0	0	4	31	12	8
Lung disease	1	2	7	11	1	3	3	23	12	8
Liver disease	1	2	9	15	3	10	3	23	16	10
High blood pressure	3	6	13	21	4	13	4	31	24	15
Ulcers	6	12	9	15	7	23	4	31	26	17

*For each condition and respondent type, the non-response rate is the number of respondents who answered "Don't know"/number of respondents; includes both first and second respondents.

RESULTS

We contacted potential first respondents for 124 (84%) of the 147 decedents. Of those contacted, 96 (77%) participated in the study. During the interview, the first respondents identified 73 potential second respondents (50% of the 147 decedents). Fifty-nine of these 73 individuals (81%) participated in the study. Among decedents for whom a proxy was interviewed, 60% of the proxies were listed as the informant on the death certificate.

Decedents with and without a proxy did not differ by gender or age, or by causes of death other than cancer and chronic lung disease. Based on underlying and contributory cause of death information on the death certificate, decedents for whom we interviewed a proxy respondent were more likely to have cancer and less likely to have chronic lung disease than decedents for whom we did not interview a proxy.

The 96 first respondents included 44 spouses (46%), 34 children and parents (35%), 14 siblings (15%), and 4 other relatives and friends (4%). The 59 second respondents included 6 spouses (10%), 27 children and parents (46%), 17 siblings (28%), and 9 other relatives and friends (15%). The median time between death and interview for all respondents was 12 years. As the time since death increased, the type of proxy respondent changed. For decedents with a more recent date of death, the respondent was more likely to be a child or sibling than a spouse. Sixty-six percent of respondents were female. The median length of time for which respondents had known the decedents was 41 years.

The ability of proxy respondents to provide information depended on their relationship to the decedent and to the level of detail requested. Non-response rates for all respondents combined ranged from 5% for cancer and heart disease to 17% for ulcers (Table 2). For each medical condition, the rate of non-response was lowest among spouses, intermediate among children, parents, and siblings, and highest among other relatives and friends (Table 2). Spouses had the lowest non-response rates for questions about the type of heart and liver disease (8% of 24 and 0% of 9 respondents, respectively). Children and parents had the lowest non-response rates for questions about the type of cancer and lung disease (4% of 23 and 7% of 14 respondents, respectively). There was no consistent relationship between the time since death and the rate of non-response.

The agreement between two different proxy respondents for the same decedent varied by condition. Table 3 shows the kappa values for pairs in which neither respondent answered "don't know". Excellent agreement ($\kappa > 0.75$) was obtained for ulcers, cancer, and diabetes. Good agreement ($0.40 \leqslant \kappa < 0.75$) was obtained

Table 3. Agreement between two proxy respondents about the occurrence of medical conditions in deceased study subjects

	No. pairs*	Kappa
Ulcers	41	0.91
Cancer	55	0.89
Diabetes	49	0.84
Lung disease	49	0.77
Heart disease	53	0.58
Stroke	49	0.55
Liver disease	47	0.45
High blood pressure	42	0.47

*For each condition, this column shows the number of pairs for which both respondents answered "yes" or "no". Pairs in which one or both of the respondents answered "don't know" were excluded from the kappa calculations.

Table 4. Relationship between time since death and success in retrieving medical records for medical conditions reported by proxy respondents*

	Obtained record			Did not obtain record		
	N	Mean†	(SD)	N	Mean	(SD)
Cancer	22	10.2	(7.2)	13	15.9	(9.2)‡
Heart disease	21	8.6	(5.5)	19	12.5	(9.0)
Stroke	8	9.8	(7.0)	4	12.6	(12.6)
Ulcers	8	7.1	(5.5)	9	11.9	(7.0)
High blood pressure	7	10.1	(9.8)	22	8.9	(7.7)
Liver disease	6	4.6	(6.2)	9	11.1	(9.6)
Lung disease	4	4.7	(3.6)	13	11.6	(9.6)
Diabetes	4	7.0	(11.9)	7	3.8	(2.8)
Any condition	29	16.0	(8.6)	55	9.4	(6.7)

*Includes only data from the first respondent when there were two respondents.
†Number of years between death and proxy respondent interview.
‡$p \leq 0.05$.

for heart disease, high blood pressure, stroke, and liver disease.

Our ability to obtain medical records depended on both the proxy respondent and on the provider. Respondents knew the name of the diagnosing physician for 117 (43%) of 270 reported conditions and the name of the hospital for 193 (89%) of 217 reported hospitalizations. Recall of physician name was greatest for spouses (60% of reported conditions), intermediate for children and parents (40%), and other relatives and friends (36%), and lowest for siblings (18%). The pattern was similar for recall of hospital name, except that sibling recall equalled that of other relatives and friends. Two hundred and fifty-two unique medical providers were identified. Medical records were obtained from 126 (50%) of these providers. The remaining records were not retrieved due to our inability to locate the provider (15%), the provider's inability to locate the records (31%), and the provider's refusal to participate (4%).

Success in obtaining medical records depended on the number of years since death, the relationship between the proxy respondent and the decedent, and the type of medical condition. For all conditions except diabetes and high blood pressure, the mean time since death was shorter for decedents whose records were obtained than for decedents whose records were not obtained (Table 4). The proportion of medical records obtained for decedents with spouse respondents was greater than that obtained for decedents with other types of respondents for all medical conditions except diabetes and lung disease (Table 5). Overall, medical record retrieval rates ranged from a low of 24% for high blood pressure cases to a high of 86% for cancer cases (Table 5).

Sixty percent or more of the medical records were retrieved for four conditions: cancer, heart disease, stroke, and liver disease. The PPV of the proxy respondent information for these conditions was 93, 81, 78, and 60%, respectively. For cancer and heart disease, there were no differences between the PPV of spouses and other respondents. Small numbers precluded analysis by type of respondent for the other conditions.

DISCUSSION

This study examined the extent to which different types of proxy respondents can provide accurate information about the medical history of deceased individuals in epidemiologic studies. Despite intensive efforts to locate proxy respondents, we were successful in obtaining interview data for only two-thirds of the deceased study subjects. Other studies similar to ours that have sought proxy respondents for deceased study subjects have reported obtaining proxy interviews for 59% [14] and 70% [17] of decedents. It is conceivable that our ability to locate and interview a proxy respondent was influenced by

Table 5. Proportion of medical records obtained by type of respondent and medical condition*

	Spouse respondent	Other respondent	Total
Stroke	80.0 (5)†	57.0 (7)	66.7 (12)
Cancer	84.6 (13)	50.0 (22)	62.9 (35)
Heart disease	60.9 (23)	41.2 (17)	52.5 (40)
Liver disease	88.9 (9)	16.7 (6)	60.0 (15)
Ulcers	55.6 (9)	37.5 (8)	47.1 (17)
High blood pressure	29.4 (17)	16.7 (12)	24.1 (29)
Diabetes	37.5 (8)	33.3 (3)	26.4 (11)
Lung disease	22.2 (9)	25.0 (8)	23.5 (17)

*Includes only data from the first respondent when there were two respondents.
†Number of records requested in parentheses.

the closeness of their relationship to the decedent. If those with a close relationship are more knowledgeable about the decedent's medical history, then our findings may be biased towards fewer non-responses, better aggreement between paired respondents, and higher rates of confirmation when comparing proxy information with medical records.

Spouses in this study were better able to provide information about a wider range of medical conditions than were other types of proxy respondents. Farrow and Samet [12] and Pickle *et al.* [18] also had greater success in obtaining most types of information from spouses, although in the latter study siblings more frequently provided answers to questions about early life events than spouses. Magaziner *et al.* [19] reported that siblings also answered questions about items requiring judgments about health and functional status more often than spouses. It is difficult to judge whether more information is better than less, if this information comes from respondents who feel compelled to guess or make up answers when they really do not know. Obtaining a second, independent source of information may clarify, or further confuse, the picture.

Enterline and Capt [11] found that spouses agreed with self-respondents for a wide range of medical conditions. Rocca *et al.* [5] found that for some medical conditions we examined in this study (high blood pressure, heart disease, stroke, lung disease, and ulcers) spouses agreed more often with self-respondents than did other types of proxy respondents [12]. In the present study, respondents agreed most frequently with each other about ulcers, cancer, diabetes, and lung disease. For the two most frequently reported conditions, cancer and heart disease, agreement was excellent and good, respectively. For the two least frequently reported conditions, diabetes and stroke, agreement was excellent and good, respectively. This pattern does not suggest any relationship between agreement and prevalence or severity of the condition. These results, however, may have been affected by the exclusion of a high percentage (approximately 30%) of respondent pairs in which at least one of the pair answered "don't know" for the questions about high blood pressure and ulcers.

We examined validity of proxy respondent information by calculating the positive predictive value. The predictive value for a positive cancer history reported to us by the first proxy respondent compared to the medical record was 92.3%. This is consistent with the values of 95.0 and 94.2% calculated from data reported by Love *et al.* [13] and Steenland and Schnorr [14], respectively. The predictive value for other chronic medical conditions was lower than that for cancer. In a recent review [20], Harlow and Linet noted that medical records confirmed 36–70% of self-reported chronic illness. In our study, the success rate for retrieving medical records and the accuracy of proxy respondent reports was better for life-threatening conditions, such as cancer, heart disease, stroke, lung disease, and diabetes, than for conditions perceived to be less disabling, such as high blood pressure and ulcers.

Several sources of error potentially influence our findings regarding validity of proxy information. In a large percentage of cases, proxy respondents were not able to identify a medical provider for us to constact and, among those providers who were identified and contacted, 41% did not furnish us with the requested records. The records we obtained were more likely to come from decedents who had died within 10 years of the proxy interview or for whom the proxy respondent was the spouse. This is likely to result in overestimates of the true positive predictive value. Another potential problem arises from the quality of the medical records themselves. When information recorded in medical records differs from that obtained from interviews, it is not possible to determine which is more accurate or complete. When the medical records are incomplete, but proxy respondents provide accurate information, the predictive value will be underestimated.

Our findings underscore some of the difficulties in using proxy respondent information and in comparing interview information with medical records. When proxy respondents are essential to a study, we suggest that investigators examine several decedent–proxy respondent relationships to determine who is the best respondent, and seek a second, independent source of information whenever possible. Obtaining this information from medical records, however, may not be feasible for individuals who died more than 10 years prior to the study.

Acknowledgements—The New Jersey Department of Health conducted this study under contract to the National Institute for Occupational Safety and Health with funding from the Agency for Toxic Substances and Disease Registry. Kathleen O'Leary was instrumental in facilitating the study. The authors also thank Jolene Schottelkotte, Brent Tomp-

kins, Steve Adams, and Becky Swart for their assistance in records and data management and the National Center for Health Statistics for reviewing medical records.

REFERENCES

1. Rogot E, Reid DD. The validity of data from next-of-kin in studies of mortality among migrants. **Int J Epidemiol** 1975; 4: 51–54.
2. Kolonel LN, Hirohata T, Nomura AMY. Adequacy of survey data collected from substitute respondents. **Am J Epidemiol** 1977; 106: 476–484.
3. Marshall J, Priore R, Haughey B, Rzepka T, Graham S. Spouse–subject interviews and the reliability of diet studies. **Am J Epidemiol** 1980; 112: 675–683.
4. Herrmann N. Retrospective information from questionnaires. I. Comparability of primary respondents and their next-of-kin. **Am J Epidemiol** 1985; 121: 937–947.
5. Rocca WA, Walter A, Fratiglioni L *et al.* The use of surrogate respondents to obtain questionnaire data in case–control studies of neurologic diseases. **J Chron Dis** 186; 39: 907–912.
6. Metzner HL, Lamphiear DE, Thompson FE, Oh MS, Hawthorne VM. Comparison of surrogate and subject reports of dietary practices, smoking habits and weight among married couples in the Tecumseh diet methodology study. **J Clin Epidemiol** 1989; 42: 367–375.
7. McLaughlin JK, Mandel JS, Mehl ES, Blot WJ. Comparison of next-of-kin with self-respondents regarding questions on cigarette, coffee, and alcohol consumption. **Epidemiology** 1990; 1: 408–412.
8. Pershagen G, Axelson O. A validation of questionnaire information on occupational exposure and smoking. **Scand J Work Environ Health** 1982; 8: 24–28.
9. Coggon D, Pippard EC, Acheson ED. Accuracy of occupational histories obtained from wives. **Br J Ind Med** 1985; 42: 563–564.
10. Shalat SL, Christiani DC, Baker Jr EL. Accuracy of work history obtained from a spouse. **Scand J Work Environ Health** 1987; 13: 67–69.
11. Enterline PE, Capt KG. A validation of information provided by household respondents in health surveys. **Am J Public Health** 1959; 49: 205–212.
12. Farrow DC, Samet JM. Comparability of information provided by elderly cancer patients and surrogates regarding health and functional status, social network, and life events. **Epidemiology** 1990; 1: 370–376.
13. Love RR, Evans AM, Josten DM. The accuracy of patient reports of a family history of cancer. **J Chron Dis** 1985; 38: 289–293.
14. Steenland K, Schnorr T. Availability and accuracy of cancer and smoking data obtained from next-of-kin for decedents in a retrospective cohort study. **J Occup Med** 1988; 30: 348–353.
15. Sweeney MH, Fingerhut MA, Connally LB *et al.* Progress of the NIOSH cross-sectional medical study of workers exposed to chemicals contaminated with 2,3,7,8-TCDD. **Chemosphere** 1989; 19: 973–977.
16. Fleiss JL. **Statistical Methods for Rates and Proportions.** New York: John Wiley; 1981.
17. 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.
18. Pickle LW, Brown LM, Blot WJ. Information available from surrogate respondents in case–control interview studies. **Am J Epidemiol** 1983; 118: 99–108.
19. Magaziner J, Simonsick EM, Kashner TM, Hebel JR. Patient–proxy response comparability on measures of patient health and functional status. **J Clin Epidemiol** 1988; 41: 1065–1074.
20. Harlow SD, Linet MS. Agreement between questionnaire data and medical records. **Am J Epidemiol** 1989; 129: 233–248.