Retrospective Smoking History Data Collection for Deceased Workers: Completeness and Accuracy of Surrogate Reports

Jessica Graus Woo, MHSA
Susan M. Pinney, PhD

Few studies have considered the completeness and accuracy of smoking histories from surrogates for deceased workers. We surveyed 68 surrogates for uranium plant workers who completed smoking histories before their deaths. Completeness: 96% of surrogates answered definitively (yes or no) about overall smoking status (eg, used any tobacco type), and 79 to 100% of those answered definitively about specific tobacco types. Of those reporting positive history, 50% to 67% answered detailed questions. Accuracy: Surrogates were accurate about overall smoking status (Kappa (κ) = 0.75) compared with index self-report, but tended to underreport (P = 0.10). Spouses and nonspouses performed similarly. Accuracy by tobacco type was moderate (κ = 0.21 to 0.56), with cigar smoking underreported. Surrogates for cigarette smokers underreported duration and amount smoked. Surrogates for deceased workers can report accurately on overall smoking status, but underreport other smoking habits. (J Occup Environ Med. *2002;44: 915–923*)

The use of surrogate reporting in epidemiological studies is often necessary, especially when the subject of interest, the index, is unavailable or deceased. As part of a larger study of former workers at the now-closed Fernald Feed Materials Processing Center (FMPC) near Cincinnati, Ohio, we investigated the accuracy and completeness of surrogate reports of smoking history for a cohort of deceased workers.

The Fernald FMPC was a Department of Energy (DOE) uranium ore processing center in operation from 1952 until 1989. Approximately 7300 individuals worked in the plant for at least 3 months during the 37-year operation. It is one of only two DOE sites that processed enriched ore from the Belgian Congo, resulting in high radon emissions from the K-65 silos on the property.

We began a study of the former Fernald employee population in 1999 to model radon exposure estimates and collect smoking histories for the Fernald workers, in preparation for an epidemiologic study of the health effects of radon exposure in this population. Accurate smoking history information is critical in such a study, as the effects of chronic low-dose radon exposure are unclear, particularly for nonsmokers.[1,2] Studies of miners have suggested that the interaction between smoking and radon exposure on lung cancer risk is between additive and multiplicative in scale.[3–5] In addition, the interaction between smoking and radon exposure may depend on age,[5,6] timing

From the University of Cincinnati, Department of Environmental Health, Division of Epidemiology and Biostatistics.

Address correspondence to: Jessica Graus Woo, MHSA, PhD Candidate, University of Cincinnati, Department of Environmental Health, Division of Epidemiology and Biostatistics, P.O. Box 670056, Cincinnati, OH 45267-0056; jagwoo@yahoo.com.

DOI: 10.1097/01.jom.0000031925.77204.ca

of radon versus cigarette exposure[7] or duration of cigarette smoking.[8]

While there is sufficient radiation data to model worker exposures in various jobs, only scant and nonquantitative smoking history data is available in plant medical records. In addition, many of the former workers have died, requiring much of the smoking history data to be collected through surrogates, often spouses. To date, we have collected 442 surrogate-reported smoking histories for deceased workers from this overall cohort in addition to 2174 self-reported smoking histories collected through large-scale mailings and the Fernald Workers' Medical Monitoring Program, described below. This data collection is ongoing, and will eventually include as many complete smoking histories as possible from the total cohort of ~7300.

The question of surrogate accuracy for smoking history data has been considered since at least 1975.[9] Few studies have focused on surrogates for deceased workers,[10,11] however, especially for those who may also have experienced harmful occupational exposures. Most other studies of smoking history reporting have focused exclusively on cigarette smoking. While cigarette smoking is likely the most important single source of tobacco exposure, other types of tobacco may also be important determinants of the health effects of radon and other occupational exposures.[12] Therefore, this study attempts to determine surrogate accuracy not only for cigarettes, but also for cigar and pipe smoking and the use of chewing tobacco.

Materials and Methods

In 1994, former workers of the Fernald plant agreed to a settlement in class action litigation against the US DOE. One aspect of the settlement provided a lifetime voluntary Fernald Workers' Medical Monitoring Program, which includes yearly health questionnaires and evaluations to determine any adverse health effects that may be related to their work exposure to radiation from Fernald. Information in the first health questionnaire includes a detailed smoking history, including lifetime smoking history, amount, duration and quitting habits for cigarettes, cigars, pipes, and chewing tobacco. Potential subjects (indexes) were 108 former Fernald employees who had completed an initial health questionnaire for this program, but who had subsequently died. Surrogate respondents were identified from up to three sources of information in the index's medical chart or elsewhere: (1) name of the index's spouse provided on the initial health questionnaire, (2) name of "survivor" on the application for the litigation settlement compensation, or (3) names and addresses of up to "three people who do not live with you but will always know where you are" provided on the initial health questionnaire. Spouses were chosen preferentially as potential surrogates, with adult children of the index or other relatives chosen when information about the spouse was missing, the spouse could not be contacted, or the spouse refused to participate.

Surrogates were contacted with a mailed information packet, including a letter of introduction, an informed consent form, the 6-page questionnaire, a half-sheet encouraging participation even if the index did not smoke, and a return addressed stamped envelope. The half-sheet was included in the mailing after feedback to a pilot questionnaire suggested that participants were less likely to return smoking history questionnaires about nonsmokers. Surrogates with available phone numbers were contacted two weeks after the mailing to encourage participation, and a second mailing was sent to non-respondents within a month of the first mailing. All surrogates included in the analysis signed a statement of informed consent, a component of the study protocol approved by the University of Cincinnati Institutional Review Board.

Questionnaires for surrogates were designed to match the questions posed to index subjects as part of the Medical Monitoring Program. Surrogates were asked to provide demographic and relationship information about themselves and the index (eg, familial relationship, length of time they lived in the same house, year the index died). They were also asked about the index's smoking history, including whether the index ever used any tobacco product, whether they used cigarettes, cigars, pipes, or chewing tobacco in particular, and detailed information about each type of tobacco, including the duration that the index used that type of tobacco and the average amount that the person smoked per day or week. The questionnaire was designed with a "skip pattern" design; if the index did not use a particular type of tobacco, the detailed questions would be skipped. Educated guessing was encouraged, but a "Don't know/Cannot guess" option was also included for each question. Questionnaires were personalized with the deceased index's name and personal pronouns throughout the questionnaire. Data from both the self and surrogate questionnaires was independently entered by two individuals into a SAS data set and compared for consistency. Discrepancies in data entry were resolved by one of the researchers (JGW). All data analysis was conducted using SAS software (SAS Institute, Cary, NC).

Positive "overall smoking status" was determined if the index met one or more of the following criteria in his or her lifetime: (1) smoked at least 400 cigarettes (20 packs), (2) smoked at least one cigar per week for a year or longer, (3) smoked at least one pipe per week for a year or longer, or (4) used chewing tobacco at least once a week for a year or longer. In addition, smoking history for any tobacco type was separately determined using the same criteria.

TABLE 1
Characteristics of Study Population Compared with Non-Responders and Other Fernald Cohorts*

	Study Cohort		Fernald Cohorts		
	Indexes with Surrogate Responders	Indexes without Surrogate Responders	Deceased with Smoking History	Living with Smoking History	Total Fernald Cohort
N	68	40	442	2174**	7402**
Information source	Self/Surrogate	Self	Surrogate	Self	Fernald records
Gender (% male)	96%	93%	96%	82%	84%
Tenure at Fernald (years [SD])	14 (11.7)	11.7 (10.8)	13.5 (9.2)	6.4 (8.3)	5.9 (8.0)
Pay category (% hourly)	60%	70%	70%	44%	47%
Time between death and survey mailing (years [SD])	3.4 (1.2)	3.2 (1.1)	16 (9.3)	N/A	N/A
Smoker (%)	79%	81%	76%	58%	N/A
Cigarette smoker (% of smokers)	94%	94%	93%	91%	N/A

* When both surrogate and self-report available, self-reported data included in this table.
** Includes a small number of sub-contractors to the Fernald site (up to 150 in the total Fernald cohort).

TABLE 2
Self-Reported Index Smoking History Characteristics

Characteristic	n/N Responding	%
Positive overall smoking status	52/66	79%
Positive cigarette status	49/52	94%
Positive cigar status	14/49	29%
Positive pipe status	14/49	29%
Positive chewing tobacco status	5/52	10%
Used only one type of tobacco	30/52	58%
Used two types of tobacco	14/52	27%
Used three types of tobacco	8/52	15%
Quit smoking prior to completing questionnaire	34/44	77%

Some editing was conducted such that index or surrogate questionnaire responses to duration and amount questions were considered when assigning "positive smoking history" status, particularly for those who left the overall or tobacco-type specific smoking status questions blank. Other internally inconsistent results were maintained as reported.

Agreement between index and surrogate reports was determined by the Kappa statistic, percent agreement, and McNemar's test of asymmetry for categorical variables and paired Student's *t* tests for continuous variables. A two-sided *P*-value ≤ 0.05 was considered to be statistically significant.

Results

Response Rate and Representativeness

Of the 108 questionnaires sent to surrogates, 68 were returned, for a response rate of 63%. Seventy-nine percent of surrogates were spouses of the deceased index, reflecting our preferential emphasis on spousal surrogates. Index subjects included 65 men (96%) and 3 women (4%).

This study focuses on the accuracy of responses for a small proportion of the total Fernald population. Our validity study cohort ($n = 108$), however, is representative of the entire deceased cohort with surrogate-reported smoking history ($n = 442$) in terms of gender, tenure and pay category at Fernald, and reported smoking habits (Table 1). The study cohort is somewhat different from living workers from whom we have collected smoking history data ($n = 2174$) and the total Fernald cohort ($n = 7402$), in that tenure at Fernald is shorter and the proportions of males and smokers are smaller in the living and total cohorts than in any of the deceased cohorts. This may reflect true differences between the Fernald employees during the earlier years (eg, predominantly male and employed at Fernald for longer terms), which make up the majority of the deceased cohorts, and those employed later (eg, higher percentage female and short-term employees), which comprise much of the living cohort.

Within the potential study cohort, the 68 indexes whose surrogates responded are similar to the 40 indexes of surrogate non-responders in terms of gender, tenure and pay category at Fernald, interval between the index's death and questionnaire mailing, and smoking habits.

Self-Reported Data

Two indexes did not provide information on overall smoking status or sufficient detailed smoking history, and were excluded from the analysis.

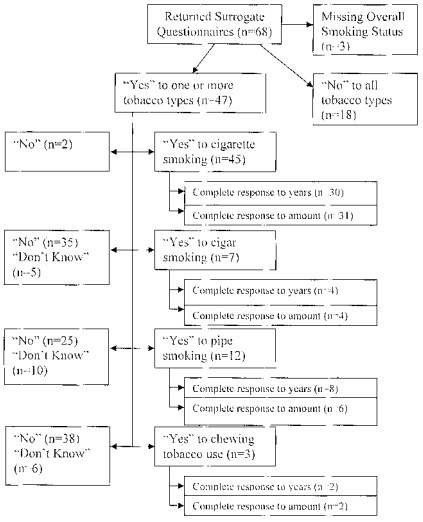

Fig. 1. Surrogate responses by smoking history question.

Seventy-nine percent of indexes (52 of 66) self-reported a positive overall smoking history, indicating that they met the criteria for using one or more tobacco types in their lifetime (Table 2). Of these, the vast majority (94%) smoked cigarettes, while no greater than 29% used cigars, pipes, or chewing tobacco. Most smokers (58%) used only one type of tobacco (usually cigarettes), but 42% of smokers used either two or three types of tobacco in their lifetime, often combining cigarette smoking with either cigar or pipe smoking or both. Seventy-seven percent of indexes (34 of 44 reporting) reported that they had quit using at least one type of tobacco before completing the questionnaire.

Surrogate Completeness

Although 68 surrogates completed questionnaires, the response rate varied from question to question, as summarized in Fig. 1. Three surrogates did not provide information on overall lifetime smoking status or sufficient detailed smoking history, and were therefore removed from the analysis. Of the remaining 65 surrogates, forty-seven indicated that the index used one or more tobacco types, while 18 indicated a negative lifetime smoking history.

The 47 surrogates reporting a positive overall smoking history were asked to respond to each type of smoking in particular (eg, cigarettes, cigars, pipes, and chewing tobacco). While all of the surrogates were able to provide a definitive response (either yes or no) about cigarette smoking status, only 89%, 79%, and 87% of surrogates could provide a definitive response (either yes or no) about cigar smoking, pipe smoking, and chewing tobacco use, respectively. The 11 to 21% of "missing" responses was overwhelmingly "Don't Know/Cannot Guess" with very few blanks. This suggests that surrogates may be relatively more aware of their index's cigarette smoking status than other types of tobacco use, even when they could report definitively on overall smoking status.

The number of non-responses was even greater when we asked detailed questions about the duration and amount smoked, by tobacco type. Among the 45 surrogates indicating positive cigarette smoking history, only 30 (67%) could provide information about the number of years smoked, and only 31 (69%) provided information about average cigarettes smoked per day. We found similar patterns for all tobacco types, with only 57% of surrogates for cigar smokers, 50 to 66% of surrogates for pipe smokers, and 66% of surrogates for chewing tobacco users responding to detailed questions. The vast majority of the nonresponses were "Don't Know/Cannot Guess," suggesting insufficient knowledge of the index's smoking habits rather than an unwillingness to answer.

Surrogate Accuracy: Overall Lifetime Smoking Status

After removing all pairs in which either the index or proxy did not respond to the overall smoking history question ($n = 5$), we analyzed 63 pairs of surveys for surrogate accuracy. Seventy-one percent of surrogates indicated the index smoked, versus 78% of index self-reports, resulting in a simple Kappa (κ) value of 0.75 (95% confidence interval [CI]: 0.56 to 0.94), with 90% exact agreement. McNemar's test of symmetry had a P-value of 0.10, indicating a trend toward underreporting among all surrogates taken together.

Spouses and non-spouses performed similarly in reporting on index overall smoking status, with κ of 0.74 and 0.80, respectively (test of equal κ, $P = 0.77$). Subgroup analyses of spouses and non-spouses revealed neither group was biased toward overreporting or underreporting overall smoking status ($P = 0.18$ and 0.32, respectively). Percent agreement was quite high within each group, with 90% of spouses and 92% of non-spouses agreeing on the index's lifetime smoking status.

Recency of the index's death appears to affect surrogate reporting accuracy on overall smoking history. Surrogates reporting on indexes who had died within the previous three years ($n = 38$, mean recall = 2.5 years) tended to provide accurate results ($\kappa = 0.80$, 95% CI: 0.60 to 1.00), but marginally underreported smoking status ($p = 0.08$). Surrogates reporting for those who had died three to six years previously ($n = 25$, mean recall = 4.5 years) appeared somewhat less accurate ($\kappa = 0.65$, 95% CI: 0.29 to 1.00) but were more balanced in their misreporting ($p = 0.56$). Statistically, the two Kappa coefficients are not distinguishable ($P = 0.47$).

Surrogate Accuracy: Smoking Status by Tobacco Type and Quitting Patterns

For each type of smoking, we included only those pairs in which both the index and surrogate responded definitively (eg, yes or no) about that type of tobacco use. Surrogate accuracy in reporting smoking status varied by the type of tobacco reported. Table 3 shows that agreement, shown as "+/+" or "−/−" in the table, favors either high preva-

TABLE 3
Index/Surrogate Report of Smoking Status, Overall and by Type of Tobacco

Category	*n**	+/+**	+/−	−/+	−/−
Overall smoking status	63	44	5	1	13
	(100)	(70)	(8)	(2)	(21)
Cigarette status	49	42	3	1	3
	(78)	(86)	(8)	(2)	(6)
Cigar status	42	3	8	3	28
	(67)	(7)	(19)	(7)	(44)
Pipe status	34	6	4	5	19
	(54)	(18)	(12)	(15)	(56)
Chewing tobacco status	47	2	3	1	41
	(75)	(4)	(6)	(2)	(87)

* *n* indicates total index-surrogate pairs with both providing definitive responses (either yes or no) for this question. Top number, count; bottom number, percent. Percent in this column calculated as number of definitive responses in each category over number definitive responses to overall smoking status (63).

** For all four remaining columns, +, positive smoking history; −, negative smoking history. First symbol represents index self-report and second symbol represents surrogate response. Top number, count; bottom number, percent. Percent calculated as number under the column heading (eg, '+/+') over number of definitive responses in the category (eg, cigarettes).

TABLE 4
Agreement between Index- and Surrogate-Reported Smoking Status, Overall and by Tobacco Type*

Category	*n*	% Agree	Kappa (95% CI)	McNemar *P*-value
Overall smoking status	63	90%	0.75 (0.56–0.94)	0.10
Cigarette status	49	92%	0.56 (0.17–0.94)	0.32
Cigar status	42	74%	0.21 (−0.12–0.53)	0.13
Pipe status	34	74%	0.38 (0.05–0.71)	0.74
Chewing tobacco status	47	91%	0.46 (0.01–0.90)	0.32

* *n* indicates total index-surrogate pairs with both providing definitive responses (either yes or no) for this question, excluding all missing data. CI, confidence interval.

lence (eg, cigarettes) or low prevalence (eg, chewing tobacco) tobacco types. Agreement was lower and less uniform for tobacco types that are used with intermediate frequency (eg, cigars and pipes).

Percent agreement between surrogates and indexes was best for cigarette smoking status (Table 4). Forty-five of 49 reporting surrogate-index pairs (92%) agreed about cigarette smoking status, either positive or negative, resulting in a moderately high Kappa score ($\kappa = 0.56$). Surrogates did not appear consistently to over-report or underreport cigarette-smoking status. In contrast, indexes and surrogates did not agree well on cigar smoking ($\kappa = 0.21$). In addition, there was a trend for surrogates of smokers to underreport cigar smoking when they were incorrect ($P = 0.13$). Surrogates achieved a similarly low agreement for pipe smoking behavior, but with a somewhat higher Kappa and no noticeable asymmetry. Chewing tobacco use was rare in this population, with most of the responses affirming non-use, leading to somewhat better agreement in status, with no asymmetry in response.

Surrogates were moderately accurate with respect to index quitting patterns. It is important to note that the time frame considered in the two questionnaires was different, with indexes answering about their quitting by the time they completed the survey (generally when they enrolled in the Medical Monitoring Program), and their surrogate answering about quitting by the time the index died. Of those reporting ($n = 44$), 89% of the surrogates reported that the index had quit smoking, while only 77% of indexes had reported quitting ($\kappa = 0.61$, 95% CI: 0.31 to 0.91). McNemar's test reveals significant asymmetry in responses, with surrogates likely to over-report quitting ($P = 0.025$), with quitting patterns among cigarette smokers driving this asymmetry ($P = 0.05$). Because of the differences in timing between the two reports, some indexes may have quit smoking after entering the Medical Monitoring Program, leading to the (potentially correct) higher surrogate report of indexes' quitting patterns.

Surrogate Accuracy: Detailed Smoking History

Surrogate data obtained about detailed smoking history of indexes was limited. Of the 42 surrogate-index pairs who both reported positive cigarette smoking status ("+/+" in Table 2), only 29 (69%) both provided information about the number of years smoked and 28 (67%) both reported the number of cigarettes smoked per day. Surrogates reported 3.6 fewer years of cigarette smoking ($P = 0.08$) and 2.9 fewer cigarettes per day ($P = 0.13$) than index self-reports (Table 5). Due to small numbers, we cannot report results for cigars ($n = 2$), pipes ($n = 5$), and chewing tobacco use ($n = 1$) with respect to accuracy about amount or duration of use.

We further investigated the impact of reporting on duration and amount of use on the final calculation of pack-years of smoking, a common

TABLE 5
Index- and Surrogate-Reported Amount and Duration of Cigarette Use and Calculated Pack Years*

		Mean (SD)		Mean Difference	
	n	Index	Surrogate	(SE)	*P*-value**
Cigarettes smoked/day	28	20.2 (8.2)	17.3 (9.0)	2.9 (1.8)	0.13
Years smoked	29	36.2 (14.8)	32.7 (17.0)	3.6 (2.0)	0.08
Pack years	23	38.4 (22.0)	34.9 (28.6)	3.6 (4.4)	0.43

* *n* indicates total index-surrogate pairs reporting a positive cigarette smoking status and both providing responses for this question. SD, standard deviation; SE, standard error of the mean difference. Pack years calculated as {(cigarettes smoked/day)/20}/(years smoked), and is defined as the number of years the index smoked a pack of 20 cigarettes a day.

** *P*-value from paired *t* test.

epidemiologic measurement of tobacco exposure. Twenty-three of the 42 index-surrogate pairs (55%) both provided sufficient information to calculate pack-years smoked. The variance of pack year estimates for both indexes and surrogates was quite large, precluding any statistical difference between them. However, the pattern of underreporting among surrogates appears consistent with the above findings.

Discussion

This study examined both the completeness and accuracy of surrogate reporting of smoking history data for a deceased worker population. With respect to completeness, we found that surrogates were progressively less likely to answer more detailed questions. Nearly all of the surrogates in our study were able to answer definitively about the index's overall lifetime smoking status. Among those who reported a positive overall status (eg, used one or more tobacco types), all were able to respond about cigarette smoking in particular, but between 11 and 21% of surrogates were unable to answer about cigar smoking, pipe smoking or chewing tobacco use. "Nonresponse" rates for more detailed smoking questions about duration and amount of use by tobacco type ranged from 33% to 50%. Only three other studies have previously investigated non-response rates for surrogate-provided smoking histories. Pickle's[13] study of 2606 next-of-kin of deceased or disabled lung cancer patients found non-response rates among spouses of between 16% and 44%, with the "detailed smoking history" most frequently left blank. Swanson[14] also investigated surrogate completeness among cancer patients or their next-of-kin ($n = 107$). They report that all patients and surrogates provided information about cigarette smoking status, but that surrogates accounted for between 56% and 86% of the "unknown" responses for detailed smoking history questions. In contrast to our study, however, Nelson et al[15] report that surrogate non-response was less than 10% for detailed cigarette smoking questions, and that spouses had lower non-response rate than other surrogates. While these results are not directly comparable to our study, it appears that surrogates frequently leave detailed questions blank.

With regard to accuracy of reported responses, our study confirms that surrogates can report accurately on an index's overall lifetime smoking history. We found that 90% of surrogates and indexes agreed about the index's lifetime use of any tobacco product. Recall interval, however, may create subtle shifts in recall accuracy among surrogates for deceased subjects. Importantly, agreement between indexes and surrogates was similar for the two different recall periods studied (<3 years and 3 to 6 years), but differed in their likelihood to underreport smoking status. This effect of time since death on reporting accuracy has not previously been investigated, and our study suggests that recall intervals up to 6 years may introduce only subtle shifts in surrogate accuracy.

Some studies have found that spouses perform better as surrogates than non-spouses,[11,16,17] while others have not.[15,18] Two studies have found that spouses report smoking status and duration of smoking with greater sensitivity and specificity than nonspouses,[16] particularly colleagues.[11] Nelson et al[15] investigated detailed smoking histories for 283 control-surrogate pairs and 68 case-surrogate pairs in a study of subarachnoid hemorrhage. Their study noted high agreement on smoking status for all surrogate categories ($\kappa = 0.74$ to 0.95) and little difference between spousal and nonspousal reports on detailed questions. Lyon et al[18] reported that spouses even performed somewhat worse than nonspouses on categorizing ever smoking, and were more likely to underestimate pack-years. Our study found that spousal and nonspousal surrogates provided similar agreement with their indexes, with $\kappa = 0.74$ and $\kappa = 0.80$, respectively, not a significant difference. In addition, although all surrogates taken together tended to underestimate the index's overall smoking status, our subgroup analysis did not implicate either spouses or non-spouses as particular contributors to this underreporting.

We investigated differences in reporting accuracy for different tobacco types. Agreement on smoking status was 92% for cigarettes and 91% for chewing tobacco use, the former predominantly confirming use and the latter confirming nonuse. Our surrogates' agreement on cigarette smoking status compares well with previous reports of 85% to 100% agreement for "current" or

"ever" cigarette smoking,[9,10,15–31] and is nearly identical to surrogates for another cohort of deceased workers (92% agreement).[11] Agreement in our study, however, was only 74% for both cigar and pipe smoking status. This is lower than reported agreement for cigar and pipe smoking of approximately 80% in both living and deceased cohorts,[22,26] with one study reporting relatively high agreement (>85%) for all tobacco types in a living cohort.[20] In addition, we report a trend toward underreporting of cigar smoking status when the index and surrogate disagreed. In our study, agreement by type of tobacco was a function of the prevalence of that tobacco's use, with higher accuracy for tobacco types with either high or low prevalence of use. This is in contrast to Boyle and Brann's[16] finding that lower prevalence smoking habits are reported with low sensitivity.

Poor surrogate-index agreement on non-cigarette smoking may materially affect retrospective studies, particularly with lung cancer as an outcome. A large recent European study found that cigar or pipe smoking confers significantly increased odds of developing lung cancer (odds ratio [OR] = 9.0 for cigars/cigarillos and 7.9 for pipe smoking) compared with never smokers, and each showed a dose-dependent relationship.[32] Our finding of lower agreement on cigar and pipe smoking and underreported cigar smoking status may mask true relationships between these exposures and lung cancer risk in retrospective studies of deceased workers, if surrogate data is used.

Detailed smoking history provided by surrogates has often been criticized as less accurate,[11,15,16,18,31,33] and our own study confirms this assertion. Other studies have found agreement between indexes and surrogates on cigarettes/day typically to be low, with exact agreement estimated at 36% when measured continuously[19] and from 29 to 84% when agreement is within five cigarettes/day.[19,26,27] Use of four to eight smoking categories rather than continuous measures of cigarettes/day improves percent exact agreement (range: 46% to 87%), but the Kappa statistics remain notably low (κ = 0.33 to 0.60).[9,11,23,31,34] Surrogate reports on total years smoked appear to be somewhat more stable and accurate (κ = 0.84 to 0.92).[15,22]

Our study focused on surrogate under- or over-reporting of duration and amount smoked, and the effects on pack-year calculations. Underreporting by surrogates on detailed smoking history is relatively common,[11,15,16,31] but not universal,[9,22,35] and may be influenced by the index's disease status.[33] Nelson et al report that surrogates underreport years of cigarette smoking by 1.68 years (95% CI: −3.24, −0.12), and tend to underreport cigarettes/day (−1.96, 95% CI: −4.2, 0.28), with a resulting insignificant 2 pack-year difference.[15] In contrast, paired responses on years smoked or cigarettes/day between 66 deceased lung cancer patients and their wives were similar.[22] In our study, surrogates on average reported 3.6 fewer years of cigarette smoking and 2.9 fewer cigarettes/day than indexes, both marginally significant findings. This underreporting, however, did not result in significantly lower estimates of pack-years, consistent with previous reports.[15]

There are several potential reasons for our surrogate inaccuracy. First, for our population of mostly-male working indexes, the surrogate was often a female spouse who may not have been aware of her husband's smoking habits at work. Interviews with living former Fernald workers indicate that many workers smoked at work during their breaks or at lunch. Second, many of the workers' families were aware of the potential occupational risks involved in working at Fernald. Families of deceased Fernald workers have shared with us concerns that their family members' illnesses were associated with their tenure at Fernald. It is possible that family members downplayed the index's smoking habits in line with their beliefs that occupational exposures, rather than lifestyle choices, adversely impacted their loved ones' health. Although one small study of surrogates for deceased smelter workers (n = 14) found that surrogate accuracy was unaffected by knowledge about the risk of smoking combined with arsenic exposure at the plant,[10] it seems likely that surrogate knowledge of occupational risks would play a significant role in studying deceased occupational cohorts. Finally, it is possible that some of the spouses married the index later in life, and would therefore be less likely to know the index's early smoking history. Our spousal surrogates (n = 54), however, reported living with the index an average of 44.8 (SD = 13.3) years, and only 4% of our entire sample lived with the index fewer than 16 years. Therefore, we do not believe remarriage late in life to be a major source of bias.

This study provides some insight into the interplay between completeness (ie, willingness to respond) and accuracy of surrogates. One hypothesis is that people who are less sure do not respond, leaving only a more accurate subset of responders. Therefore, one might expect that for general questions, most people would be confident in their answer and would respond, which our study reports. One would also expect, however, that with more detailed questions, fewer people would respond (which we found), and that those people who did respond would be more accurate (which we did not find). Thus, this hypothesis does not hold in our study population. This interaction between confidence and accuracy, however, has not previously been reported and should be explored further in relation to surrogate reporting.

This study is one of only three that has considered a deceased index in an occupational context, and the other studies are inadequate to answer important questions in occupational epidemiology. One includes a cohort of 118 welders, but the recall

interval was short (<2 years),[11] and the other is a small study of only 14 smelter workers which used unverified plant medical records as the reference.[10] Most studies of non-occupational cohorts have focused on cigarette smoking histories obtained from living indexes and surrogates, often obtained concurrently. The applicability of these studies to retrospective cohorts of occupational populations may, therefore, be limited. In particular, these studies do not often address the recall interval implicit in collecting data on deceased individuals, or the fact that indexes may be aware that their answers will be compared with a surrogate's.

In conclusion, we have found that surrogates for a deceased working population agree sufficiently with indexes about the index's lifetime smoking status, but are less accurate regarding the duration or amount of smoking. Spouses and non-spouses perform similarly in reporting smoking status. Agreement about smoking status does vary by tobacco type, however, and is subtly influenced by the length of recall time after the index's death. Thus, this study supports the usefulness of querying surrogates about their family member's smoking status for retrospective occupational epidemiologic studies, but data on non-cigarette smoking status, smoking duration or amount smoked should be interpreted with caution.

Acknowledgments

Grateful acknowledgment is extended to James Nasuta, Ronald Freyberg, Stacy Poe, and Ellen Galbraith for significant work on interviewing workers, developing our database and administering the Fernald Workers' Medical Monitoring Program.

Funding support for this work was provided by the National Institutes of Occupational Safety and Health (Grant No. R01/CCR515748).

References

1. Lubin JH, Boice JD, Edling C, et al. Lung cancer in radon-exposed miners and estimation of risk from indoor exposure. *J Natl Cancer Inst.* 1995;87:817–827.
2. Steenland K, Thun M. Interaction between tobacco smoking and occupational exposures in the causation of lung cancer. *J Occup Med.* 1986;28:110–118.
3. Hornung RW, Deddens JA, Roscoe RJ. Modifiers of lung cancer risk in uranium miners from the Colorado Plateau. *Health Phys.* 1998;74:12–21.
4. Hornung RW, Deddens J, Roscoe R. Modifiers of exposure-response estimates for lung cancer among miners exposed to radon progeny. *Environ Health Perspect.* 1995;103(Suppl 2):49–53.
5. Moolgavkar SH, Luebeck EG, Krewski D, Zielinski JM. Radon, cigarette smoke, and lung cancer: a re-analysis of the Colorado Plateau uranium miners' data. *Epidemiology.* 1993;4:204–217.
6. Steenland K. Age specific interactions between smoking and radon among United States uranium miners. *Occup Environ Med.* 1994;51:192–194.
7. Thomas D, Pogoda J, Langholz B, Mack W. Temporal modifiers of the radon-smoking interaction. *Health Phys.* 1994; 66:257–262.
8. Band P, Feldstein M, Saccomanno G, Watson L, King G. Potentiation of cigarette smoking and radiation: evidence from a sputum cytology survey among uranium miners and controls. *Cancer.* 1980;45:1273–1277.
9. Rogot E, Reid DD. The validity of data from next-of-kin in studies of mortality among migrants. *Intl J Epidemiol.* 1975; 4:51–54.
10. Pershagen G, Axelson O. A validation of questionnaire information on occupational exposure and smoking. *Scand J Work, Environ Health.* 1982;8:24–28.
11. Hansen KS. Validity of occupational exposure and smoking data obtained from surviving spouses and colleagues. *Am J Indust Med.* 1996;30:392–397.
12. Qiao YL, Taylor PR, Yao SX, et al. Relation of radon exposure and tobacco use to lung cancer among tin miners in Yunnan Province, China. *Am J Indust Med.* 1989;16:511–521.
13. Pickle LW, Brown LM, Blot WJ. Information available from surrogate respondents in case-control interview studies. *Am J Epidemiol.* 1983;118:99–108.
14. Swanson GM, Schwartz AG, Brown KL. Population-based occupational cancer incidence surveillance. Utilization of the telephone interview. *J Occup Med.* 1985; 27:439–444.
15. Nelson LM, Longstreth WT, Koepsell TD, Checkoway H, van Belle G. Completeness and accuracy of interview data from proxy respondents: demographic, medical, and life-style factors. *Epidemiology.* 1994;5:204–217.
16. Boyle CA, Brann EA. Proxy respondents and the validity of occupational and other exposure data. The Selected Cancers Cooperative Study Group. *Am J Epidemiol.* 1992;136:712–721.
17. Gilpin EA, Pierce JP, Cavin SW et al. Estimates of population smoking prevalence: self- vs. proxy reports of smoking status. *Am J Public Health.* 1994;84: 1576–1579.
18. Lyon JL, Egger MJ, Robison LM, French TK, Gao R. Misclassification of exposure in a case-control study: the effects of different types of exposure and different proxy respondents in a study of pancreatic cancer. *Epidemiology.* 1992;3:223–231.
19. Kolonel LN, Hirohata T, Nomura AM. Adequacy of survey data collected from substitute respondents. *Am J Epidemiol.* 1977;106:476–484.
20. Herrmann N. Retrospective information from questionnaires. I. Comparability of primary respondents and their next-of-kin. *Am J Epidemiol.* 1985;121: 937–947.
21. Greenberg RS, Liff JM, Gregory HR, Brockman JE. The use of interviews with surrogate respondents in a case-control study of oral cancer. *Yale J Biol Med.* 1986;59:497–504.
22. Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol.* 1986;123:481–489.
23. Rocca WA, Fratiglioni L, Bracco L, Pedone D, Groppi C, Schoenberg BS. The use of surrogate respondents to obtain questionnaire data in case-control studies of neurologic diseases. *J Chronic Dis.* 1986;39:907–912.
24. McLaughlin JK, Dietz MS, Mehl ES, Blot WJ. Reliability of surrogate information on cigarette smoking by type of informant. *Am J Epidemiol.* 1987;126: 144–146.
25. Machlin SR, Kleinman JC, Madans JH. Validity of mortality analysis based on retrospective smoking information. *Stat Med.* 1989;8:997–1009.
26. 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.
27. Thorogood M, Vessey M. The reliability of surrogate information about oral contraceptive use, smoking, height and weight collected from men about their wives. *Contraception.* 1989;39:401–408.
28. Hatch MC, Misra D, Kabat GC, Kartzmer S. Proxy respondents in repro-

ductive research: a comparison of self- and partner-reported data. *Am J Epidemiol.* 1991;133:826–831.
29. Semchuk KM, Love EJ. Effects of agricultural work and other proxy-derived case-control data on Parkinson's disease risk estimates. *Am J Epidemiol.* 1995; 141:747–754.
30. Hyland A, Cummings KM, Lynn WR, Corle D, Giffen CA. Effect of proxy-reported smoking status on population estimates of smoking prevalence. *Am J Epidemiol.* 1997;145:746–751.
31. Passaro KT, Noss J, Savitz DA, Little RE. Agreement between self and partner reports of paternal drinking and smoking. The ALSPAC Study Team. Avon Longitudinal Study of Pregnancy and Childhood. *Intl J Epidemiol.* 1997;26:315–320.
32. Boffetta P, Pershagen G, Jockel KH, et al. Cigar and pipe smoking and lung cancer risk: a multicenter study from Europe. *J Natl Cancer Inst.* 1999;91: 697–701.
33. Weiss A, Fletcher AE, Palmer AJ, Nicholl CG, Bulpitt CJ. Use of surrogate respondents in studies of stroke and dementia. *J Clin Epidemiol.* 1996;49:1187–1194.
34. Halabi S, Zurayk H, Awaida R, Darwish M, Saab B. Reliability and validity of self and proxy reporting of morbidity data: a case study from Beirut, Lebanon. *Intl J Epidemiol.* 1992;21:607–612.
35. Barnett T, O'Loughlin J, Paradis G, Renaud L. Reliability of proxy reports of parental smoking by elementary school-children. *Ann Epidemiol.* 1997;7:396–399.