MAYO CLINIC

ORIGINAL ARTICLE

Accuracy of Smoking Status Reporting: Proxy Information in a Rapidly Fatal Cancer Setting

Maria A. Stevens, MA, MPH; Kari G. Rabe, MS; Ben Boursi, MD; Aarti Kolluri, BS; Dhruv P. Singh, MD; William R. Bamlet, MS; and Gloria M. Petersen, PhD

Abstract

Objective: To assess whether patients and relatives can serve as reliable proxy reporters of other family members' cigarette-smoking history.

Patients and Methods: Two samples (325 patients, 707 relatives) were identified from the Mayo Clinic Biospecimen Resource for Pancreas Research, enrolled from November, 6, 2000, to March 15, 2018. Smoking-history data, including categorical (ever/never) and quantitative (packs per day and years smoked) smoking measures, were obtained from self-completed questionnaires by patients and relatives. Relative reports were compared with patient reports on self; patient reports were compared with relative reports on self.

Results: Overall, spouses and first-degree relatives (FDRs) were accurate (94.5%) when reporting patient ever smoking; spouse reports were 98.6% sensitive and 97.7% accurate. Accuracy of patient reports was 97.8% for spouse smoking and 85.5% for FDR smoking; accuracy varied by relationship of FDR. When not concordant, patients generally over-reported daily packs smoked by relatives and under-reported years smoked. Within a 25% agreement range, spouse reports about patients' daily packs smoked was 46.7%, and years smoked was 69.6%, whereas FDRs were 50% and 64.6%, respectively. When not concordant, relatives generally over-reported daily packs smoked by patients, but no consistent pattern was observed of over- or under-reporting years smoked by patients.

Conclusions: Patients and relatives can be reliable proxies for smoking history (ever/never) in their family members, especially spouses. An accurate reporting of smoking status will help physicians to better gauge performance status and family smoking exposures to inform disease management.

 ■ Mayo Clin Proc Inn Qual Out 2020;4(6):801-809

Pancreatic cancer (PC) ranks as the third leading cause of cancer deaths among men and women in the United States, with an estimated 57,600 incident cases in 2020.[1] It is a rapidly fatal cancer; for 92% of patients diagnosed with localized, regional, or metastatic disease, the 5-year survival is 37.4%, 12.4%, and 2.9%, respectively.[2] The consensus in the literature is that cigarette smoking is attributable to approximately 25% of PC cases,[3] with reported relative risks ranging from 0.3 to 5.4 in current smokers and 0.6 to 2.4 in former smokers.[4] The significant variability of relative risks that is seen among studies is due in part to participation bias, particularly for patients with a rapidly fatal cancer, such as PC, as the self-respondent is often deceased or unavailable (eg, overburdened by intensive medical procedures) to provide self-reported or biochemical smoking information. As such, researchers often resort to smoking information provided by a proxy or next of kin.

Proxy reports can increase case representation and statistical power, allowing researchers to quantify the influence of smoking on rarer and lethal diseases such as PC more comprehensively. Previous studies have shown that proxies can be accurate reporters of smoking status for patients with colon,[5,6] breast,[7] lung,[8,9] renal,[10] and other cancers[11-13] but less accurate when reporting quantitative smoking history. Reporting accuracy differed, based on type of proxy and measurement method used.[8,10] For example, some studies demonstrated good agreement between

From the Department of Health Sciences Research, Mayo Clinic, Rochester, MN (M.A.S., K.G.R., A.K., D.P.S., W.R.B., G.M.P.); Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC (M.A.S.); Department of Oncology, Sheba Medical Center, Tel Hashomer, Israel (B.B.); and Department of Medicine, Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN (D.P.S., G.M.P.).

spouses for "ever smoked"[8] and increased response rates in spouses and children for smoking status and history compared with other proxies.[7,9] However, other studies found increased response rates and good rates of exact agreement across proxy reporters for smoking status[10,12] but higher rates of exact agreement in spouses for detailed smoking history.[10] The discrepancies in findings hamper researchers from determining which specific proxies and measures most accurately represent cases' smoking histories.

To our knowledge, only 1 study measured proxy reporting of smoking for self-respondents with PC, and measures were limited to response rates of smoking status in which 44% of spouses were unable to report detailed smoking histories.[12] Further, no studies to date have assessed the accuracy and concordance of patients' reporting smoking history about their first-degree relatives (FDR) and spouses. The latter is important if patient proxy reports provide more comprehensive insight to clinicians and researchers about family history of smoking as a risk factor (eg, environmental exposure or clustering patterns of smoking) in the setting of familial pancreatic cancer—those with 2 or more FDRs affected with PC.[14] The purpose of this study was to determine whether FDRs (children, parents, and siblings) and spouses can serve as accurate and concordant proxy reporters of cigarette smoking histories for patients with PC and, reciprocally, to quantify the accuracy and concordance of patient proxy reports about FDRs and spouses on the same smoking measures including over- and underestimation of smoking measures. Our findings will ultimately help to inform researchers who wish to study the influence of cigarette smoking in rapidly fatal cancers and provide estimates of exposure accuracy for clinicians who may use the information in disease management.

METHODS

Study Design and Procedures

Using a cross-sectional design, this study analyzed questionnaire data obtained from the Mayo Clinic Biospecimen Resource for Pancreas Research, a registry that uses prospective ultra-rapid case ascertainment to recruit patients with PC and, subsequently, their relatives, with an estimated 80% rate of participation.[15] Relatives and patients enrolled between November 6, 2000, and March 15, 2018, were asked to complete a Risk-Factor Questionnaire (RFQ) that included questions about demographic information, family history, and cigarette-smoking history on both the respondent and their relatives. Of the 2826 consecutively enrolled patients in the registry who had completed RFQs, we invited family members to also participate, focusing on families that contained at least 2 relatives with PC or PC and melanoma. This resulted in 325 families in which RFQ data were completed by both patients and at least 1 family member who were then included for analysis in this study. Among the 325 patients, 59.7% reported family histories of PC, 6.8% reported family histories of melanoma, and 33.5% had no family histories of PC. We observed no differences in sex, race, ethnicity, smoking status, or cancer stage among the 2501 patients who were not included in the study and our final study sample of patients; however, there was a slightly younger mean age at diagnosis of PC among the study-sample patients (64.4±10.1 standard deviation [SD]) compared with those not in the study (65.9±10.4 SD, $P<.01$). Relatives typically completed their own RFQs within 6 months of patient self-completion of their RFQs.

The following questions were asked of each patient and his or her relatives about themselves: Have you ever smoked cigarettes (more than 100 in your lifetime? (No/Yes); if "Yes," how old were you when you first started smoking cigarettes? __ years old. On average, how many cigarettes do/did you smoke per day? (20 cigarettes per pack) ___. Do you currently smoke cigarettes (No/Yes); if "No," Did you quit within the last year? No/Yes. At what age did you quit? __ years old. How many total years have you smoked cigarettes on a regular basis?

Each respondent was also asked to provide a detailed family history and asked, relative by relative, if the person ever smoked and, if so, daily packs smoked and years smoked. This study was approved by the Mayo Clinic Institutional Review Board, and subjects agreed to participate in this research by providing informed consent.

Mayo Clin Proc Inn Qual Out ■ December 2020;4(6):801-809 ■ https://doi.org/10.1016/j.mayocpiqo.2020.07.010
www.mcpiqojournal.org

Study Subjects

Two samples were constructed: Sample I (patient reporting on relatives) and Sample II (relatives reporting on patient). Subjects met inclusion criteria for Sample I if they had participated in the Mayo Clinic Biospecimen Resource for Pancreas Research, had been previously diagnosed with PC, had completed an RFQ, and had at least 1 paired RFQ completed by their corresponding FDR or spouse. Sample I included 325 patients with PC—consecutively enrolled—who reported about their relatives (consisting of 31 parents, 333 siblings, 218 children, and 145 spouses). Subjects met inclusion criteria for Sample II if they had participated in the Mayo Clinic Biospecimen Resource for Pancreas Research and had completed an RFQ with at least 1 paired RFQ completed by their corresponding affected patient. Sample II included 707 relatives ascertained through a patient with PC, including 31 parents, 313 siblings, 218 children, and 145 spouses, who reported about the affected patient.

Statistical Analysis

Responses were compared with self-reported smoking history completed by the corresponding patient or relative(s). Owing to the unavailability of biochemical measures of smoking—often used as a gold standard in smoking cessation studies[16]—self-reported cigarette smoking history was treated as "truth." Specificity, sensitivity, and accuracy analyses were performed to compare patient and relative reports with self-reports for the question, "Have you ever smoked cigarettes (more than 100 in your lifetime)?" Current and former smokers were considered "ever smokers." Sensitivity was defined as the extent in which the proxy report agreed with the subject who ever smoked. Specificity was defined as the extent in which the proxy report agreed with the subject as to nonsmokers. Binomial 95% confidence intervals (CIs) were computed. To be included in the analyses, both the patient and relative in a given pair both had to declare responses to each question (ie, each had to respond with "yes" or "no" to smoking, and for packs and years smoked, both had to respond "yes" to smoking, and both had to list packs and years smoked; missing responses from either respondent would lower the potential sample counts). Concordance of responses were calculated for whether the patient or relative responded within 10% and 25% ranges of the self-reported quantitative values for packs smoked daily and years smoked. For example, if a patient self-reported that he or she smoked 1.0 pack per day for 10 years, the relative must respond within 0.9 to 1.1 and 0.75 to 1.25 packs and 9 to 11 and 7.5 to 12.5 years, respectively, to achieve a concordant response for packs and years smoked. We determined these cut-points to be close enough to the self-report (truth) to be concordant while overcoming misclassification bias if researchers were to employ similar proxy reports in follow-up studies. This was further extended to comparing patients' and relatives' over- and under-reporting of daily packs smoked and years smoked. All analyses were conducted using SAS v9.4 software (SAS Institute, Cary, North Carolina).

RESULTS

Characteristics of Samples

Demographic characteristics for Samples I and II are shown in Table 1. For Sample I, there were 176 (54.2%) men and 149 (45.9%) women. Mean age was 64.8 years (SD=10.1), and the sample was 99.7% white and 99.4% non-Hispanic/non-Latino. In Sample I, 179 (55.6%) patients reported that they had smoked cigarettes; the mean number of packs smoked daily was 1.0 (SD=0.6), and mean number of years smoked was 23.8 (SD=14.6). For Sample II, there were 261 (36.9%) men and 446 (63.1%) women, and mean age was 66.6 years (SD=9.8). In this sample, 327 (45.0%) relatives reported that they had smoked cigarettes, with a mean of 0.8 (SD=0.6) number of packs smoked daily and a mean of 18.7 (SD=14.4) number of years smoked.

Sample I: Patients Reporting About Relatives. Sensitivity, specificity, and accuracy of ever or never smoking are presented in Table 2. Overall, accuracy of patient reports about their spouses and FDRs was 87.9%. Patients demonstrated a sensitivity of 74.3% when reporting about their relatives.

TABLE 1. Characteristics of Respondents in Study Samples

Respondent characteristics	Sample I: Patients N=325	Sample II: Relatives N=707
Relationship to patient, n (%)		
Parents	-	31 (4.4)
Children	-	218 (30.9)
Spouses	-	145 (20.5)
Siblings	-	313 (44.2)
Number of relatives		
Mean (SD)	2.19 (1.70)	-
Median (minimum-maximum)	1 (1-11)	-
Sex, n (%)		
Female	149 (45.9)	446 (63.1)
Male	176 (54.2)	261(36.9)
Age, mean (SD)	64.8 (10.1)	66.6 (9.8)
Race, n (%)		
White	324 (99.7)	699 (99.2)
Non-white	1 (0.3)	6 (0.8)
Ethnicity, n (%)		
Non-Hispanic/non-Latino	312 (99.4)	665 (99.7)
Hispanic/Latino	2 (0.6)	2 (0.3)
Educational attainment, n (%)		
< High school	11 (3.4)	25 (3.5)
High school graduate	88 (27.1)	144 (20.4)
Some college	75 (23.1)	188 (26.6)
College graduate	73 (22.5)	214 (30.3)
Postgraduate	78 (24.0)	135 (19.1)
Ever smoker, n (%)	179 (55.6)	327 (45.0)
Cigarette packs smoked daily, mean (SD)	1.0 (0.6)	0.8 (0.6)
Years smoked, mean (SD)	23.8 (14.6)	18.7 (14.4)

SD = standard deviation.

Sensitivity was higher when patients reported about their spouses (98.3%) compared with their FDR(s) (68.7%), although rates varied by type of FDR. Patients were accurate (98.5%) when reporting that relatives were nonsmokers.

Table 3 shows results for range of agreement of patient reports about relatives with regard to daily packs smoked and years smoked. Overall, the concordance for reports on spouses and relatives within 10% and 25% ranges of agreement were, respectively, 29.3% and 40.1% for daily packs smoked, and 27% and 52.5%, respectively, for years smoked. Responses differed by reports on spouses and type of FDR. A higher proportion of patients reported within 10% and 25% for packs smoked daily (34.0% and 48.0%) and years smoked (34.8% and 63.0 %) for their spouses compared with FDRs for packs smoked daily (27.4% and 36.8%), and for years smoked (23.2% and 47.4%). Of the patient's FDRs, a higher percentage of patients reported within 10% for packs smoked for their children's self-reports (33.3%) and years smoked for their siblings' self-reports (25.4%). Concordance within these ranges was further examined by patients who over- or underestimated daily packs smoked and years smoked or were within range. Table 4 shows that patients were in modest agreement for daily packs smoked by relatives: from 29.3% to 40.1% in the 10% and 25% ranges and 27% to 52.5%, respectively, for years smoked.

Mayo Clin Proc Inn Qual Out ■ December 2020;4(6):801-809 ■ https://doi.org/10.1016/j.mayocpiqo.2020.07.010
www.mcpiqojournal.org

TABLE 2. Sensitivity, Specificity, and Accuracy of Reported Cigarette Smoking

Sample I: Patient reports about relatives (N=325)

		Sensitivity		Specificity		Accuracy	
Relationship	N	Percent (95% CI)[a]	P/S	Percent (95% CI)[a]	P/S[c]	Percent (95% CI)[a]	P/S
FDRs and spouses	688	74.3 (69.0-79.2)	223/300	98.5 (96.7-99.4)	382/388	87.9 (85.3-90.3)	605/688
Spouses	137	98.3 (90.6-100.0)	56/57	97.5 (91.3- 99.7)	78/80	97.8 (93.7-99.5)	134/137
FDRs	551	68.7 (62.5-74.5)	167/243	98.7 (96.7-99.6)	304/308	85.5 (82.3-88.3)	471/551
Parents	30	84.6 (54.6-98.1)	11/13	100.0 (80.5-100.0)	17/17	93.3 (77.9- 99.2)	28/30
Siblings	308	71.1 (63.6-77.8)	118/166	98.6 (95.0-99.8)	140/142	83.8 (79.2-87.7)	258/308
Children	213	59.4 (46.4-71.5)	38/64	98.7 (95.2-99.8)	147/149	86.9 (81.6- 91.1)	185/213

Sample II: Relative reports about patients (N=632)

		Sensitivity		Specificity		Accuracy	
Relationship	N	Percent (95% CI)[a]	P/S	Percent (95% CI)[a]	P/S	Percent (95% CI)[a]	P/S
FDRs and spouses	632	93.4 (90.4, 95.8)	342/366	95.9 (92.7, 97.9)	255/266	94.5 (92.4, 96.1)	597/632
Spouses	132	98.7 (92.7, 100.0)	73/74	96.6 (88.1, 99.6)	56/58	97.7 (93.5, 99.5)	129/132
FDRs	500	92.1 (88.4, 94.9)	269/292	95.7 (92.0, 98.0)	199/208	93.6 (91.1, 95.6)	468/500
Parents	28	91.7 (61.5, 99.8)	11/12	87.5 (61.7, 98.5)	14/16	89.3 (71.8, 97.7)	25/28
Siblings	280	92.6 (87.4, 96.1)	150/162	97.5 (92.7, 99.5)	115/118	94.6 (92.01.3, 97.0)	265/280
Children	192	91.5 (85.0, 95.9)	108/118	94.6 (86.7, 98.5)	70/74	92.7 (88.1, 96.0)	178/192

[a]95% CI = exact binomial 95% confidence interval.

FDRs = first-degree relatives; P = proxy report of ever or never smoking; S = self-report of ever or never smoking.

Patients generally over-reported daily packs smoked by relatives and under-reported years smoked.

Sample II: Relatives Reporting About Patients. Sensitivity, specificity, and accuracy for ever or never smoking by relatives about their index patient are presented in Table 2. Spouses and FDRs demonstrated higher overall accuracy of 94.5%, with sensitivity of 93.4% and 95.9% if the patient was a nonsmoker. Reporting about the patient, both spouse and FDR reporting had high accuracy (97.7% and 93.6%), respectively. Sensitivity and specificity of reporting across the subgroups was consistently high.

Concordance of spouse and FDR reporting about daily packs smoked and years smoked by the index patient are shown in Table 3. Within 10% and 25% range of agreement, overall concordance for daily packs smoked were 36.9% and 49%, respectively, and, for years smoked, 40.2% and 66.3%, respectively. A higher proportion of spouses (41.7%) reported within 10% of the patient's response compared with FDRs (34.9%) for packs smoked daily. Of the patients' FDRs, their children were more likely to report within 10% of the index patients' self-reports for packs and years smoked: 40.5% and 45.3%, respectively. Table 4 shows that spouses and FDRs were in modest agreement for daily packs smoked by patients: from 36.9% to 49.0% in the 10% and 25% ranges, and 40.2% to 66.3%, respectively, for years smoked. Relatives generally over-reported daily packs smoked by patients but were split between over- and under-reporting years smoked.

DISCUSSION

We measured the accuracy and concordance of reporting cigarette smoking using a cross-sectional study design through questionnaire data obtained from patients with PC and relatives who participated in the Mayo Clinic Biospecimen Resource for Pancreas Research. We calculated measures of accuracy for the dichotomous variable, ever vs never smoked, and concordance for the continuous variables, daily packs smoked, and years smoked. We also extended the analysis to explore over-

TABLE 3. Concordance of Reported Cigarette Smoking History Within 10% and 25% Range of Self-Report

Sample I: Patient reports about first-degree relatives and spouses						
			Estimate of agreement range			
			Within 10%		Within 25%	
	Relationship	Total N	N	%	N	%
Daily packs smoked	FDRs and spouses	167	49	29.3	67	40.1
	Spouses	50	17	34.0	24	48.0
	FDRs	117	32	27.4	43	36.8
	Parents	9	1	11.1	1	11.1
	Siblings	81	22	27.2	32	39.5
	Children	27	9	33.3	10	37.0
Years smoked	FDRs and spouses	141	38	27.0	74	52.5
	Spouses	46	16	34.8	29	63.0
	FDRs	95	22	23.2	45	47.4
	Parents	5	1	20.0	3	60.0
	Siblings	67	17	25.4	30	44.8
	Children	23	4	17.4	12	52.2
Sample II: First-degree relative and spouse reports about patients						
			Estimate of agreement range			
			Within 10%		Within 25%	
	Relationship	Total N	N	%	N	%
Daily packs smoked	FDRs and spouses	206	76	36.9	101	49.0
	Spouses	60	25	41.7	28	46.7
	FDRs	146	51	34.9	73	50.0
	Parents	2	0	0	0	0
	Siblings	65	19	29.2	29	44.6
	Children	79	32	40.5	44	55.7
Years smoked	FDRs and spouses	169	68	40.2	112	66.3
	Spouses	56	23	41.1	39	69.6
	FDRs	113	45	39.8	73	64.6
	Parents	3	1	33.3	1	33.3
	Siblings	46	15	32.6	29	63.0
	Children	64	29	45.3	43	67.2

FDRs = first-degree relatives.

and under-reporting these measures when there was lack of agreement. To our knowledge, this is the first study to measure the accuracy and concordance of patient and relative reporting cigarette smoking when there is a known family history of a rapidly fatal cancer such as PC.

The accuracy and concordance of responses varied based on relationship to the self-reporter and type of smoking information reported. Both patients and relatives showed relatively accurate reporting for the ever vs never smoking history; relatives demonstrated higher sensitivity and accuracy compared with patients. Spouses showed more accurate reporting across smoking measures compared with FDRs. The overall concordance was diminished on quantitative smoking measures for patients and relatives with less than 42% of spouses—who tended to show the highest concordance among relatives—reporting within 10% of the self-reporter for packs and

Mayo Clin Proc Inn Qual Out ■ December 2020;4(6):801-809 ■ https://doi.org/10.1016/j.mayocpiqo.2020.07.010
www.mcpiqojournal.org

TABLE 4. Under- and Over-Reporting of Daily Packs Smoked and Years Smoked

		Sample I: Patient reports about spouses and first-degree relatives			
		Within 10% range		Within 25% range	
		N	%	N	%
Daily packs smoked	Under-reported	43	25.8	31	18.6
	Within range	49	29.3	67	40.1
	Over-reported	75	44.9	69	41.3
	Total	167	-	167	-
Years smoked	Under-reported	61	43.3	42	29.8
	Within range	38	27.0	74	52.5
	Over-reported	42	29.8	25	17.7
	Total	141	-	141	-
		Sample II: Spouse and first-degree relative reports about patients			
		Within 10% range		Within 25% range	
		N	%	N	%
Daily packs smoked	Under-reported	47	22.8	34	16.5
	Within range	76	36.9	101	49.0
	Over-reported	83	40.3	71	34.5
	Total	206	-	206	-
Years smoked	Under-reported	51	30.2	31	18.3
	Within range	68	40.2	112	66.3
	Over-reported	50	29.6	26	15.3
	Total	169	-	169	-

years smoked. In Sample II, children's reports were comparable with those of spouses on packs and years smoked; however, concordance was still diminished, with less than half reporting within 10% of the index patients' self-reports. When not concordant, patients generally over-reported daily packs smoked by relatives and under-reported years smoked. Analogously, relatives generally over-reported daily packs smoked by patients, but no consistent pattern was observed of over- or under-reporting years smoked by patients.

Our findings are not surprising, given the superior reporting on categorical measures compared with quantitative measures observed in proxy smoking responses in other cancers. For example, Boyle et al. studied the accuracy of reporting Vietnam military service exposures related to cancer and found 98% and 88% sensitivity in spouses and other proxies when indicating whether the self-respondent was a regular smoker.[11] Similarly, Pickle et al. found that spouses, siblings, and offspring were able to provide broad levels of smoking information about the subject; however, response rates were lower for detailed smoking information.[9] In contrast, McLaughlin et al. showed high rates of exact agreement between spouses when reporting the number of nonfilter cigarettes (kappa=.53) and pack-years (kappa=.58) in cases with renal cancer;[10] however, quantitative smoking measures were dichotomized into groups that may overlook heterogeneity within smoking categories (eg, 1 to 2 packs). Consistent with our findings and the literature,[7,9] information-related bias (eg, recall bias) may influence reporting based on disease status, especially on quantitative measures, as these data are difficult to report with precision. Also, spouses may be more likely to share information, thereby improving accuracy and concordance when compared with other relatives. Further, spouses as well as children of

the affected patient may be more aware of current and previous smoking behaviors and are likely to share environmental exposures compared with siblings and parents who no longer live with the patients.

Study Limitations

There are limitations to the interpretation of these findings. First, although the Mayo Clinic Biospecimen Resource for Pancreas Research registry is dedicated in its efforts to recruit family members of affected patients, not all relatives participate, nor do they provide smoking family history on all relatives, resulting in variability of pair comparisons because of missing data. Participation bias could have inflated reporting accuracy with samples biased toward patients and relatives who knew more detailed smoking information about their relatives. Social desirability bias could have also confounded results such that respondents may have under-reported cigarette smoking to be perceived in a favorable manner. In addition, reporters could be biased by disease status (ie, recall bias) and more likely to indicate smoking about family members if they have—or there is a strong family history of—cancer known to be related to use of tobacco. Some studies have employed biochemical measures of residual carbon monoxide—found in saliva, urine, and blood—as a gold standard of smoking estimates to overcome biases associated with self-report; however, these measures are often difficult to obtain, as they require extra effort on behalf of the patient or relative to provide a biospecimen.[16] We therefore chose to use self-reports, shown to be relatively comparable with biochemical measures with a mean sensitivity of 87.5% and mean specificity of 89.2% across 26 studies.[17] It is also important to consider that the homogeneity of our samples—99.7% and 99.2% white/Caucasian for patients and relatives, respectively—may not reflect race or ethnicity of all affected patients with PC and their unaffected relatives, warranting further investigation.

Study Strengths

The major strength of our study was having knowledge of truth (self-reports), so we were able to assess sensitivity, specificity, and accuracy. We also assessed over- and under-reporting of daily packs smoked and years smoked, which has not been done systematically. To our knowledge, many studies measured inter-rater agreement[6,7,10,13,18] and rates of response and nonresponse,[7,9] whereas few studies used sensitivity and specificity.[11,15,19] When validating a proxy report against a self-report, caution should be taken when relying solely on measures of inter-rater agreement because they aggregate responses across multiple reporters. Multiple reporters may also be less logistically difficult to obtain in a clinical setting. Further, response rates fail to measure how closely the proxy aligns with the self-report.

CONCLUSIONS

Case-control studies of highly fatal diseases, such as PC, often rely on information reported from a proxy or next of kin to study the relationship between exposure and progression of disease. Otherwise, misclassification bias—when the exposure is not accurately specified—can contribute to imprecision and inaccurate interpretation of study findings. Our findings illustrate that relatives, especially spouses, are more accurate when reporting cigarette smoking status (ie, ever smoking) but less concordant when reporting quantitative cigarette smoking history (ie, years and packs smoked).

This study highlights important considerations for carrying out future research when using family history of smoking reported by patients or relatives. Researchers and physicians may reasonably rely on cigarette smoking status obtained from the spouse and FDRs when self-reported smoking status by the patient is unavailable. Our work suggests that proxy reporting of smoking status can help physicians as they gauge the performance status and family smoking exposures to inform management of disease.

ACKNOWLEDGMENTS

The authors would like to thank all of our study participants and devoted staff of the Mayo Clinic Biospecimen Resource for Pancreas Research.

Abbreviations and Acronyms: **CI** = confidence interval; **FDR** = first-degree relative; **PC** = pancreatic cancer; **RFQ** = Risk-Factor Questionnaire; **SPORE** = Specialized Program of Research Excellence; **SD** = standard deviation

Grant Support: This research was funded in, part, by Mayo Clinic Specialized Program of Research Excellence (SPORE) P50CA102701, R01CA97075, U01 CA210138; Rolfe Pancreatic Cancer Foundation; and the Pancreatic Cancer Action Network.

Potential Competing Interests: The authors report no competing interests.

Correspondence: Address to Gloria M. Petersen, PhD, Department of Health Sciences Research, Mayo Clinic, 200 First Street SW, Rochester, MN 55905 (petersen.gloria@mayo.edu; Twitter: @GMPetersen1).

ORCID

Gloria M. Petersen: https://orcid.org/0000-0003-3424-7039

REFERENCES

1. American Cancer Society. Cancer Facts and Figures. https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2020.html. Accessed May 13, 2020.
2. National Cancer Intitute. Surveillance Epidemiology and End Results Program. *Cancer Stat Facts: Pancreatic Cancer*. https://seer.cancer.gov/statfacts/html/pancreas.html. Accessed July 2, 2019.
3. Raimondi S, Maisonneuve P, Lowenfels AB. Epidemiology of pancreatic cancer: an overview. *Nat Rev Gastroenterol Hepatol*. 2009;6(12):699-708.
4. Lugo A, Peveri G, Bosetti C, et al. Strong excess risk of pancreatic cancer for low frequency and duration of cigarette smoking: A comprehensive review and meta-analysis. *Eur J Cancer*. 2018;104:117-126.
5. Herrmann N. Retrospective information from questionnaires. I. Comparability of primary respondents and their next-of-kin. *Am J Epidemiol*. 1985;121(6):937-947.
6. Kaerlev L, Lynge E, Sabroe S, et al. Reliability of data from next-of-kin: results from a case-control study of occupational and lifestyle risk factors for cancer. *Am J Ind Med*. 2003;44(3):298-303.
7. Prochazka M, Hall P, Granath F, et al. Validation of smoking history in cancer patients. *Acta Oncol*. 2008;47(6):1004-1008.
8. Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol*. 1986;123(3):481-489.
9. Pickle LW, Brown LM, Blot WJ. Information available from surrogate respondents in case-control interview studies. *Am J Epidemiol*. 1983;118(1):99-108.
10. McLaughlin JK, Mandel JS, Mehl ES, et al. Comparison of next-of-kin with self-respondents regarding questions on cigarette, coffee, and alcohol consumption. *Epidemiology*. 1990;1(5):408-412.
11. 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(6):712-721.
12. Campbell PT, Sloan M, Kreiger N. Utility of proxy versus index respondent information in a population-based case-control study of rapidly fatal cancers. *Ann Epidemiol*. 2007;17(4):253-257.
13. Sandler DP, Shore DL. Quality of data on parents' smoking and drinking provided by adult offspring. *Am J Epidemiol*. 1986;124(5):768-778.
14. Hruban RH, Petersen GM, Goggins M, et al. Familial pancreatic cancer. *Ann Oncol*. 1999;10(Suppl 4):69-73.
15. McWilliams RR, Rabe KG, Olswold C, De Andrade M, Petersen GM. Risk of malignancy in first-degree relatives of patients with pancreatic carcinoma. *Cancer*. 2005;104(2):388-394.
16. Regan S, Reid ZZ, Kalley JHK, et al. Smoking status confirmation by proxy: validation in a smoking cessation trial. *Nicotine Tobacco Res*. 2016:34-40.
17. Patrick DL, Cheadle A, Thompson DC, Diehr P, Koepsell T, Kinne S. The validity of self-reported smoking: a review and meta-analysis. *Am J Public Health*. 1994;84(7):1086-1093.
18. Woo JG, Pinney SM. Retrospective smoking history data collection for deceased workers: completeness and accuracy of surrogate reports. *J Occup Environ Med*. 2002;44(10):915-923.
19. Hansen KS. Validity of occupational exposure and smoking data obtained from surviving spouses and colleagues. *Am J Ind Med*. 1996;30(4):392-397.