

ORIGINAL ARTICLE

Validation of smoking history in cancer patients

MICHAELA PROCHAZKA¹, PER HALL², FREDRIK GRANATH³ & KAMILA CZENE²

¹The Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden, ²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden and ³Department of Medicine at Karolinska University Hospital, Clinical Epidemiology Unit, Stockholm, Sweden

Abstract

Objective. The aim of the study was to investigate the agreement between patient records and the information reported from relatives, and how the relationship and time since patient's death affected the response rate and the quality of the data. **Methods.** A questionnaire regarding smoking history was sent to next-of-kin of 270 deceased women diagnosed with breast cancer during 1958–2000 in the Stockholm County. Agreement between the reports of next-of-kin and patient records was calculated using a kappa statistics, along with its 95% confidence interval. **Results.** When information about overall smoking history from patient records and next-of-kin was compared, the kappa value was 0.83 (95% confidence interval (CI): 0.73–0.92). Using two smoking categories (≤ 15 cigarettes, > 15 cigarettes), for quantitative smoking history the kappa value was 0.46. No evidence of a trend toward under-/over-reporting among next-of-kin was found. The overall smoking agreement between medical record and next-of-kin was similar for the two median recall periods (≤ 10 years and > 10 years), with kappa value 0.80 and 0.83, respectively. **Conclusion.** Next-of-kin can provide reliable information with almost perfect agreement with patient records on lifetime smoking status, and should be considered in studies where information on smoking history is missing.

The quality of the exposure information is of utmost importance in all epidemiological studies. This information is often obtained from patient records since data from other sources are not available. When studying the risk of lung cancer information on smoking history is vital [1]. The completeness and quality of the information on tobacco use in patient records is often not optimal and sometimes even not recorded. If the patient is deceased the only available source for information is next-of-kin. The reliability and validity of next-of-kin for smoking history has been reported in several studies [2–5]. Most studies show that misclassification of overall smoking status is low and mainly concerns light and former smokers. This is in contrast to detailed smoking history where the agreement was reported to be low [2,5–8]. Some studies have found that spouses are more reliable than other relatives [9,10], a finding contradicted by others [8,11]. Few studies have focused on information from next-of-kin for deceased patients [6,8,9,12], especially for those that died decades ago [12,13].

The purpose of this study was to investigate the completeness and accuracy of the smoking history in the patient records, the agreement between patient records and the information reported from relatives, and to investigate how relationship and time since patient's death affected the response rate and the quality of the data.

Material and methods

Our study was performed within the framework of a study of lung cancer risk in women previously treated for breast cancer in the Stockholm County during the period 1958–2000 [1]. Using the Swedish Cancer Registry (SCR), we identified 192 women diagnosed with breast cancer and subsequent lung cancer in Stockholm County. Lung cancers diagnosed within 12 months of the initial breast cancer were excluded. Controls consisted of women diagnosed with breast cancer, who survived the corresponding medical date of lung cancer diagnosis. Controls were matched by age and decade of breast

cancer diagnosis. The information about smoking habit was collected from patient records at the time of breast cancer diagnosis, therefore cases and controls were considered as one group. The total number of women were 401, 306 deceased and 95 still alive. The mean age at diagnosis of breast cancer among the 401 women included in the study was 58 years (median 57 years; 1st quartile 48 years and 3rd quartile 66 years).

In the present study, the Swedish Taxation authority, Swedish Register of Population and Population Changes were used to identify next-of-kin for the 306 deceased patients. Using the Cause of Death Registry, the national registration number and name were used to control the vital status of next-of-kin. We identified next-of kin for 270 women (Figure 1), for 16 (5%) women all relatives were deceased, and for 20 (7%) women the relatives could not be identified. Selection of the relative was done in the order: spouses, children, siblings, and other relatives [8]. The lifetime smoking habits of the study participants were validated by a questionnaire.

A letter of introduction and one page questionnaire was mailed to the relative with a return stamped envelope. A second mailing was sent to non-respondents one month after the first mailing. All living patients received a letter of introduction and the same questionnaire as next-of-kin.

The validation questionnaire was designed to match the information on smoking habits abstracted in the patient records. The information included lifetime smoking history, amount, and duration of tobacco use. No additional information on study

participants was collected. Positive “overall smoking status” was determined as occasional smoking or one cigarette or more daily. In addition, smoking history of any tobacco type was separately determined using the same criteria. The questionnaire to next-of-kin was personalized with the deceased patient’s name and year of birth and death.

The information about smoking status according to next-of kin was available for 223 patients (Figure 1). All the corresponding patient records were identified and reviewed. Information about smoking habits was registered in 132 patient records and these patients were subsequently used in the validation study.

Statistical methods

Exact percentage agreement between next-of-kin and patient records was calculated by dividing the number of agreed smokers and non smokers, by the total number of participants in the validation study. To evaluate the quality of agreement between the reports of next-of-kin and patient records, we calculated kappa statistics, along with its 95% confidence interval [14]. Landis and Koch [15] have classified kappa values into three groups with respect to the degree of chance-adjusted agreement. A kappa statistics representing less than 0.40 should be taken as a poor agreement, between 0.40 and 0.75 fair to good agreement, and greater than 0.75 almost perfect agreement. The McNemar’s test of asymmetry for categorical variables was calculated and the Cochran-Armitage trend test was used to

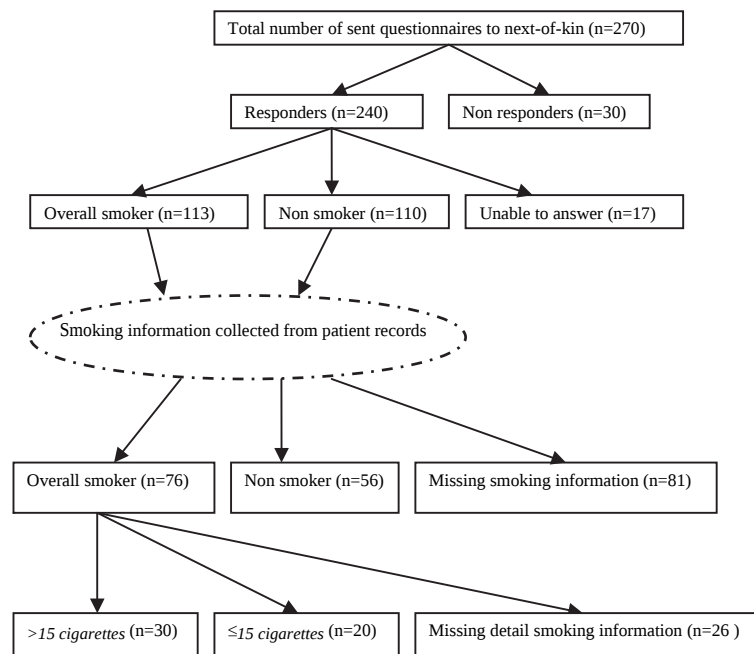


Figure 1. Flow chart of the study.

test for trends to determine the effect of time on response rate. A two-sided *p*-value was used for all analyses, and was considered to be statistically significant at $p \leq 0.05$.

Results

Of the 270 questionnaires sent to next-of-kin, 240 were returned, one answered by telephone, resulting in response rate of 89% (Figure 1). A person was counted as a responder, when the questionnaire was returned, even if no smoking information could be provided. Table I shows the proportion of responders divided by relationship to the patient. Of the 83% of next-of-kin who were able to answer about patient's overall lifetime smoking status, 69% were children of the deceased woman. The response rate from next-of-kin was not affected by the time since the woman's death (Table I), and varied between 86 to 95%. The trend test showed no evidence of changes in the response rate ($p = 0.31$). The average interval between death of the patient in the study and next-of-kin questionnaire was 14 years, and the time interval varied between 1 and 38 years.

The proportion of 270 patients that had smoking information available was 64% from the patient record and 83% from the next-of-kin, respectively. The smoking information given in patient records and questionnaires was compared for consistency (Table II). Validation of the information in the patient records and next-of-kin could be done for 132 patients with complete information from both sources. Among these patients 11 (8%) were classified differently by next-of-kin. For eight patients, the next-of-kin reported smoking of less than eight cigarettes a day or to be an occasional smoker, where in patient records the patient was reported as

a non smoker. Three women were reported as non smokers by next-of-kin, while the patients reported use of less than eight cigarettes a day. When information from patient records and next-of-kin were compared, the kappa value was 0.83 (95% CI: 0.73–0.92). McNemar's test of symmetry revealed no evidence of a trend toward under-/over-reporting among all next-of-kin (Table II). Separate analysis of the group children showed similar kappa value 0.88, and McNemar's *p*-value 0.18. When time between death of the patient in the study and the next-of-kin questionnaire was divided in median intervals, a high agreement was present in both groups. In the interval ≤ 10 years since death with kappa value 0.80 (95% CI: 0.64–0.95), and 0.83 (95% CI: 0.70–0.97) in the interval of > 10 years.

Quantitative information on smoking from both patient records and next-of-kin was available for 50 women (Table III). Two women had missing amount of smoked number of cigarettes per day, but were stated as heavy smokers in the patient records, and were therefore considered to smoke > 15 cigarettes/day. Among the 21 women who in the patient records had smoking information of ≤ 15 cigarettes a day, seven were found for whom the next-of-kin reported a greater amount. When the quantitative information from patient records and next-of-kin was compared, the kappa value was 0.46 (95% CI: 0.21–0.71). For most of the 13 women who were classified differently by next-of-kin, the difference in tobacco smoking was small and arose mainly from differences close to the cut-off point.

For comparison we analyzed also 95 living patients. The response rate for this group was 93%, two of the living patients were too ill to answer. Among the 46 living patients with smoking information from both sources, 19 patients (41%) self-reported a positive overall smoking history, and 27 patients (59%) were non smokers. The corresponding information in the patient records was 17 smokers and 26 non smokers, resulting in a kappa value 0.86 (95% CI: 0.71–1.00), with no evidence of asymmetry (McNemar *p*-value = 0.78).

Table I. Proportion of responders (%) stratified by relationship to the patient and in relation to time since death of the patient.

	Responders (proportion%)	Able to give smoking information
Relationship to the patient		
Spouse	n = 25 (93%)	n = 25
Children	n = 158 (90%)	n = 153
Sibling	n = 8 (67%)	n = 7
Other relation	n = 49 (88%)	n = 38
Total	n = 240 (89%)	n = 223
Time since death of the patient		
<5 years	n = 42 (86%)	n = 40
5 ≤ years < 10	n = 54 (89%)	n = 52
10 ≤ years < 15	n = 49 (88%)	n = 48
15 ≤ years < 20	n = 38 (95%)	n = 34
20 ≤ years < 25	n = 25 (89%)	n = 21
≥ 25 years	n = 32 (91%)	n = 28

Discussion

Epidemiological studies of etiological type rely on high quality of the exposure information, and the information is often obtained from patient records. Obviously, this information can be biased and the precision may be low due to lack of or invalid information. The aim of this study was to study the agreement between data on smoking history in patient records and the information received from next-of-kin. In addition, we investigated how

Table II. Smoking data available at time of breast cancer diagnosis in patient records compare to information from the questionnaire from next-of-kin.

		Information from Next of-kin		
		Smoker	Non smoker	Total
Patient records data	Smoker	73	3	76
	Non smoker	8	48	56
	Total	81	51	132
Overall agreement (%)		92		
kappa (95%CI)		0.83 (0.73–0.92)		
McNemar <i>p</i> -value		0.13		

relationship and time since women's death affected the response rate and the quality of the data.

Cigarette smoking is a risk factor for many cancers, but no clear association with breast cancer has been identified [16,17]. This may affect both the frequency and quality of smoking history in patient records among women diagnosed with breast cancer. In our study, the smoking history was missing in 36% patient records. Using the information from next-of-kin may also suffer from some limitations. Next-of-kin of the case may have been more likely than next-of-kin of the control to recall the smoking habit of the woman, and may have been influenced by disease status of the case. Still, in our analyses we have not seen any difference between cases and controls, with kappa value of 0.83 and 0.80 respectively. Small amounts smoked decades ago could be especially hard to remember. It is also possible that next-of-kin were not aware of the woman's smoking habits at work. The main advantage of our study is the information from next-of-kin for patients that died decades ago. We use a very good register data for the identification of next-of-kin, however this approach is only feasible when there is a register with the information of relatives.

Several studies have previously investigated response rates of next-of-kin for smoking history [8,9,18]. They reported that relatives provided a high response rate about the patient's overall lifetime smoking status. In our study only 7% of all next-of-kin were unable to provide any information on smoking habits (Figure 1). Detailed smoking history

provided by next-of-kin has often been criticized as less accurate [6,8–10]. Pickle et al. [18] has studied 2606 next-of-kin of deceased or disabled lung cancer patients, and found that 44% of the spouses could not give a detailed smoking history. In our study, among those relatives who reported a positive overall lifetime smoking status, 88% were able to respond about more detailed smoking questions. Use of two smoking categories resulted in 74% exact agreement and the kappa statistics with a fair agreement (Table III). The underreporting by next-of-kin on detailed smoking history is relatively common [6,9,10,19]. In contrast to these results, in our study there was no evidence for underreporting by next-of-kin.

The relation of the time interval between death and next-of-kin contact was previously examined [7,20]. In these studies, the time between death and the contact with next-of-kin varied between 40 weeks up to 5 years, with response rates about 90%. In our study 73% of the women died ≥ 5 years before sending the mail questionnaire to the next-of-kin (Table I). However, the response rate was approximately 90% and was not affected by the time since the patient died.

Recall interval did not influence accuracy among next-of-kin for deceased women. The overall smoking agreement between medical record and next-of-kin was similar for the two median recall periods (≤ 10 years and > 10 years), with kappa value 0.80 and 0.83, respectively. These results illustrate the accuracy of reporting both by the next-of-kin and by the patient. The effect of time since death on

Table III. Comparison of quantitative smoking information per day between patient records and next-of-kin.

		Information from Next-of-kin		
		≤ 15 cigarettes	> 15 cigarettes	Total
Patient records data	≤ 15 cigarettes	14	6	20
	> 15 cigarettes	7	23	30
	Total	21	29	50
Overall agreement (%)		74		
kappa (95%CI)		0.46 (0.21–0.71)		
McNemar <i>p</i> -value		0.78		

reporting accuracy has previously been investigated by Woo et al [8] and they suggested that recall intervals up to six years may introduce only slight shifts in next-of-kin accuracy. Our study confirms this result, and show that the recall time has little effect on the next-of-kin accuracy even decades after the patient's death.

A number of studies have validated the information of smoking status by comparing study patients with next-of-kin [6,8,19,21], and reported a good agreement. Two studies have found that spouses report overall smoking status and duration of smoking with better sensitivity and specificity than other next-of-kin [9,10]. While other studies [8,11] found that spousal and other next-of-kin provided similar agreement on smoking status (kappa value 0.74 to 0.95). In our study spouses and children had the highest response rate on overall smoking status (Table I). Furthermore, we found 94% overall agreement between information from children and information in patient records (kappa value 0.87). Interestingly, living patients provided comparable results as next-of-kin with kappa value of 0.86.

Our study revealed almost perfect agreement of the overall smoking history between patient records and next-of-kin, indicating that both sources have a high accuracy. Our results confirm the finding of other investigators that next-of-kin rarely misclassify the cigarette smoking status (smoker/non smoker) of the studied patients, but that information on quantitative smoking history is less reliable. Spouses and children perform similarly in reporting overall smoking status. Further, we show that time since death of the studied patient has no effect on the response rate or the quality of the answer. We conclude that the use of information from next-of-kin should safely be considered in studies where information on smoking history is lacking.

References

- [1] Prochazka M, Hall P, Gagliardi G, Granath F, Nilsson BN, Shields PG, et al. Ionizing radiation and tobacco use increases the risk of a subsequent lung carcinoma in women with breast cancer: Case-only design. *J Clin Oncol* 2005;23:7467.
- [2] Herrmann N. Retrospective information from questionnaires. I. Comparability of primary respondents and their next-of-kin. *Am J Epidemiol* 1985;121:937-47.
- [3] Nyberg F, Isaksson I, Harris JR, Pershagen G. Misclassification of smoking status and lung cancer risk from environmental tobacco smoke in never-smokers. *Epidemiology* 1997;8:304-9.
- [4] Lee P. Misclassification of smoking habits and passive smoking: A review of the Evidence. Berlin; 1988.
- [5] Rogot E, Reid DD. The validity of data from next-of-kin in studies of mortality among migrants. *Int J Epidemiol* 1975;4:51-4.
- [6] Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol* 1986;123:481-9.
- [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;1:408-12.
- [8] Woo JG, Pinney SM. Retrospective smoking history data collection for deceased workers: completeness and accuracy of surrogate reports. *J Occup Environ Med* 2002;44:915-23.
- [9] Hansen KS. Validity of occupational exposure and smoking data obtained from surviving spouses and colleagues. *Am J Ind Med* 1996;30:392-7.
- [10] 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;136:712-21.
- [11] Nelson LM, Longstreth WT, Jr, Koepsell TD, Checkoway H, van Belle G. Completeness and accuracy of interview data from proxy respondents: Demographic, medical, and life-style factors. *Epidemiology* 1994;5:204-17.
- [12] Pershagen G, Axelson O. A validation of questionnaire information on occupational exposure and smoking. *Scand J Work Environ Health* 1982;8:24-8.
- [13] McLaughlin JK, Dietz MS, Mehl ES, Blot WJ. Reliability of surrogate information on cigarette smoking by type of informant. *Am J Epidemiol* 1987;126:144-6.
- [14] Rigby AS. Statistical methods in epidemiology. v. Towards an understanding of the kappa coefficient. *Disabil Rehabil* 2000;22:339-44.
- [15] Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159-74.
- [16] Gammon MD, Eng SM, Teitelbaum SL, Britton JA, Kabat GC, Hatch M, et al. Environmental tobacco smoke and breast cancer incidence. *Environ Res* 2004;96:176-85.
- [17] Egan KM, Stampfer MJ, Hunter D, Hankinson S, Rosner BA, Holmes M, et al. Active and passive smoking in breast cancer: Prospective results from the Nurses' Health Study. *Epidemiology* 2002;13:138-45.
- [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] Nyberg F, Agudo A, Boffetta P, Fortes C, Gonzalez CA, G. P. A European validation study of smoking and environmental tobacco smoke exposure in nonsmoking lung cancer cases and controls. *Cancer Causes Control* 1998;2:173-82.
- [20] Poe GS, McLaughlin JK, Powell-Griner E, Parsons CR, Robinson K. The time interval between death and next-of-kin contact and its effects on response rates and data quality. *Am J Epidemiol* 1991;134:1454-62.
- [21] Kaerlev L, Lynge E, Sabroe S, Olsen J. 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;44:298-303.