J. Clin Epidemiol Vol. 49, No. 8, pp. 885–889, 1996

ELSEVIER

0895-4356/96/$15.00
PII S0895-4356(96)00026-1

Smoking Prevalence in Neighborhood and Hospital Controls: Implications for Hospital-Based Case-Control Studies

Alfredo Morabia,[1,] Steven D. Stellman,[2] and Ernst L. Wynder[2]*

[1]CLINICAL EPIDEMIOLOGY DIVISION, UNIVERSITY CANTON HOSPITAL, GENEVA SWITZERLAND, AND [2]DIVISION OF EPIDEMIOLOGY, AMERICAN HEALTH FOUNDATION, NEW YORK, NEW YORK, UNITED STATES

ABSTRACT. It is widely believed that the prevalence of smoking among hospital patients is greater than that of the general population because many conditions for which patients are hospitalized are caused by or associated with smoking, and that this increased prevalence may bias results of case-control studies of tobacco-related diseases. For this reason, many authors have suggested excluding from the control series patients hospitalized for tobacco-related illnesses. The present study investigated potential selection bias for hospital compared to neighborhood controls in studying tobacco-related diseases. The 709 cases from six U.S. cities had tobacco-related cancers or myocardial infarction. They were individually matched to one hospital control and to one neighbor. After excluding patients with tobacco-related diseases, hospital controls were less often current smokers and more often former smokers than neighborhood controls. Among male ever smokers, hospital controls tended to smoke more cigarettes per day than neighborhood controls. Compared with the U.S. population, there was an overrepresentation of smokers in neighborhood controls rather than an underrepresentation of smokers in hospital controls. Relative risk estimates varied according to type of control. Choosing between hospital and neighborhood controls in case-control studies should be dictated by criteria related to the study hypothesis, participation, or cost. In particular, exclusion of hospital controls with diseases known to be tobacco-related seems to be a successful strategy for reducing selection bias. J CLIN EPIDEMIOL 49;8:885–889, 1996.

KEY WORDS. Case-control study, hospital controls, neighborhood controls, smoking, cancer, odds ratio

INTRODUCTION

It is widely believed that the prevalence of smoking among hospital patients is greater than that of the general population because many conditions for which patients are hospitalized are caused by or associated with smoking [1]. For instance, in their classic paper Mantel and Haenszel wrote, "Hospital controls invariably yield a higher proportion of smokers for each sex than controls of comparable age drawn from the general population" [2]. When studying tobacco-related diseases, this problem is usually prevented by choosing controls who are hospitalized for diseases not related to tobacco [3]. However, the potential for selection bias due to overhospitalization of smokers and the impact of the foregoing exclusion strategy—even though frequently alluded to [1,4]—have rarely been documented.

The purpose of this study was to compare the distribution of smoking habits in two sets of age–sex–race-matched controls from a U.S. national case-control study of tobacco-related diseases with the prevalence of these smoking habits in U.S. survey data. One set of controls was taken from the same hospitals as the cases. The second group of controls was selected from among neighbors of the cases. Another objective was to examine the effect of choice of hospital or neighborhood controls on relative risk estimates of several tobacco-related cancers.

*Address reprint requests to: Alfredo Morabia, Division d'Épidémiologie Clinique, 25, rue Micheli-du-Crest, CH-1211 Geneva 14, Switzerland.
Accepted for publication on 6 December 1996.

METHODS AND MATERIALS

The data presented here are derived from an on-going hospital-based case-control study of tobacco-related cancers started in 1969 and previously described [5]. In 1980 and 1981 the study was expanded to include neighborhood controls in addition to hospital controls.

During these 2 years case diagnoses were restricted to histologically confirmed cancers of the lung, larynx, mouth, esophagus, bladder, and pancreas as well as myocardial infarction. Hospital controls were patients admitted for diseases not considered to be tobacco related. Discharge diagnoses of the hospital controls are listed in the Appendix. Although some of these discharge diagnoses would now be thought to be related directly or inversely to cigarette smoking (e.g., certain leukemias, cervical and endometrial cancer, Parkinson disease), they represent a small proportion of our controls. Controls were frequently matched to cases on age, sex, ethnic group, hospital, and room status (private, semiprivate, ward). A contractor (a subsidiary of Equifax, Inc. in Atlanta, Georgia) carried out equivalent control interview protocols in all six cities in which the study was conducted. Each month a list of cases was sent to the subcontractor; the list contained each patient's identification number, address, age, sex, and ethnicity. The subcontractor then attempted to identify a near neighbor of the case according to a rigid scheme whereby the interviewer first enumerated addresses close to that of the case, and then canvassed them in a fixed order to determine whether a matching subject resided there. Allowance was made for repeated attempts due to nonresponse and vacancies. Persons with a past history of any current index disease were not eligible. Training of the subcontractor's field staff was organized and conducted by experienced American Health Foundation (AHF) supervisory personnel, who continued to monitor performance throughout the data collection period.

The completion rate among hospital patients, defined as the number of interviews completed divided by total patients approached, was 91.6% for cases and 92.9% for controls. On the other hand, there are several ways to define a response rate among neighborhood controls. On the basis of eligible households in which screeners were able to speak to someone, the completion rate was approximately 61%. If one assumes half the refusals would have been found eligible, the completion rate would be 76%. However, no contact was made with about one-fourth of the households attempted. These hard-to-reach households may have had more active, busier occupants, who would be likely to have contributed to a lower completion rate if reached, but who would also probably be younger than the cases. If half of these are assumed eligible, then the completion rate is only 39.7%.

The final study consisted of 1024 case-control triplets, each including one case, one matched hospital control, and one matched neighborhood control. Percent distribution of triplets was as follows: New York (n = 500, 49%), Birmingham (n = 61, 6%), San Francisco (n =15, 1.5%), Chicago (n = 184, 18%), Philadelphia (n = 100, 10%), and Pittsburgh (n = 164, 16%). More than one-third of cases were interviewed in Memorial Hospital, New York. Numbers of triplets were 717 for nonblack males, 58 for black males, 238 for nonblack females, and 11 for black females. Because of small numbers we did not perform the conditional logistic regression analyses of Tables 4 and 5 for 22 male and 10 female triplets with esophageal cancer, 18 female triplets with pancreas cancer, and 40 female triplets with myocardial infarction.

The main objective of this analysis was to compare the smoking habits of these two types of controls. Data from a national survey of U.S. smoking habits published by the Centers for Disease Control (Atlanta, GA) [6] are also presented in order to locate the smoking prevalence of controls with respect to the U.S. general population. Definitions of smoking categories were similar in the case-control study and in the survey. Subjects were classified as never having smoked regularly (never smokers), current smokers, or having quit smoking for at least 1 year (ex-smokers). Because another purpose of this study was to investigate the influence of choice of controls on relative risk estimation (rather than to study new etiological hypotheses), well-established relationships were chosen as targets of most calculations, namely, the relative risks of the tobacco-related diseases with respect to cigarette smoking. Odds ratios were estimated using conditional logistic regression, via proportional hazards models [7].

RESULTS

Differences in sociodemographic characteristics between neighborhood and hospital controls were consistent in males and females (data not shown): neighborhood controls were less well educated, less likely to be Jewish or single, more often widowed, and less likely to drink coffee compared to hospital controls. There were no differences in alcohol consumption or body mass index (kg/m^2).

Tables 1 and 2 compare the smoking status of controls with a U.S. population survey that was conducted a few years later [6]. Data are shown for completely matched sets but do not include subjects who were black (except for 43- to 64-year-old males), less than 25 years old, or who were pipe/cigar smokers. Hospital controls are more similar to the U.S. population than neighborhood controls in the age groups 45–64 and 65+ years for nonblack males, the largest subgroups of the study, and for 45- to 64-year-old nonblack females. On the other hand, the smoking habits of neighborhood controls seem more representative of the corresponding sex-race specific U.S. population for 25- to 44-year-old nonblack males and for nonblack females aged 65+ years.

Table 3 shows that neighborhood controls were less often never smokers or former smokers, more frequently current smokers, and among male ever smokers, tended to smoke a smaller number of cigarettes per day compared to hospital controls.

Tables 4 and 5 compare the conditional logistic regression odds ratios (ORs) obtained using neighborhood versus those obtained with hospital controls, separately for males and females. A ratio of odds ratios (ROR) greater than unity means that the neighborhood control OR was greater than the corresponding hospital control OR. A systematic difference in smoking prevalence between the two types of controls would be expected to produce a ROR consistently above or below unity, depending on the direction of the difference in smoking prevalence between the two control groups. Situations in which the OR was statistically significant with one group of controls but not with the other are indicated with an asterisk.

In males (Table 4), there were eight RORs greater than 1.0 and six RORs smaller than 1.0. One OR (pipes/cigar and myocardial infarction) was statistically significant with hospital but not with neighborhood controls, versus three ORs statistically significant with neighborhood controls but not with hospital controls. In females (Table 5), five of six RORs were below unity. There was one situation (exsmoking and oral cavity cancer) in which the OR was statistically significant with hospital controls but not with neighborhood controls.

DISCUSSION

In our data, hospital controls were less often current smokers and more often former smokers than neighborhood controls. On the other hand, among male ever smokers, hospital controls tended to smoke more cigarettes per day than neighborhood controls. If the U.S. population is taken as a reference point [6], then there was an overrepresentation of smokers in neighborhood controls rather than an underrepresentation of smokers in hospital controls. Thus, the present results suggest that whether one uses hospital or neighborhood controls there is no hard and fast rule to predict the direction of the bias, if any, of the odds ratio for tobacco-related diseases in relation to cigarette smoking, provided that patients with tobacco-related diseases are *a priori* excluded from the control series.

The belief that hospitalized patients are more likely to smoke cigarettes is supported by studies that use population controls selected by random digit dialing as a second control group [8]. However, neighborhood controls do not necessarily conform to that belief. A review of North American case-control studies with both hospital and neighborhood controls consistently shows that smoking is more prevalent in neighbors than in hospital patients. In a Canadian breast cancer study, Stavraky and Clarke found a higher proportion of women 55 years of age or older who had ever smoked among neighborhood controls (58%) compared to hospital controls (39%), even though 8% of their hospital controls had diseases of circulatory and respiratory systems, which are often tobacco related [9]. In the same study, there were 25% current smokers among hospital controls versus 33% of neighborhood controls. Tell *et al.* compared risk

TABLE 1. Percentages of male never, former, and current smokers in cases and controls and in the U.S. population, by race and age

Race	Age[a] (years)	Smoker	Percentage Cases (n = 709)	NC[b] (n = 715)	HC[b] (n = 703)	U.S. population[c]
Nonblack	25–44		(n = 47)	(n = 47)	(n = 57)	
		Never	17	40	26	40
		Former	21	21	23	26
		Current	62	38	51	34
				p^e = NS[b]		
	45–64		(n = 430)	(n = 424)	(n = 413)	
		Never	7	29	35	26
		Former	34	29	37	44
		Current	59	42	29	30
				$p^e = 0.0002$		
	65+		(n = 179)	(n = 193)	(n = 180)	
		Never	7	32	38	34
		Former	48	37	42	50
		Current	45	31	19	16
				$p^e = 0.04$		
Black[d]			(n = 42)	(n = 34)	(n = 39)	
	45–64	Never	10	21	46	35
		Former	21	35	28	29
		Current	69	44	26	36
				$p^e = 0.06$		

[a]Data for <25 years not presented because of small sample size.
[b]NC, neighborhood controls; HC, hospital controls; NS, not significant.
[c]See Ref. 6.
[d]Data for black males aged 25–44 years and 65+ years not presented because of small sample size.
[e]p value for chi-square test (2 df) for the difference between the two types of controls.

TABLE 2. Percentages of female never, former, and current smokers in cases, controls, and in the U.S. population, by race and age

Race	Age[a] (years)	Smoker	Percentage Cases (n = 236)	NC[b] (n = 238)	HC[b] (n = 237)	U.S. population[c]
Nonblack	25–44		(n = 25)	(n = 22)	(n = 23)	
		Never	8	50	48	51
		Former	8	27	17	20
		Current	84	23	35	28
				p^d = NS[b]		
	45–64		(n = 140)	(n = 143)	(n = 140)	
		Never	20	48	57	49
		Former	14	13	19	26
		Current	66	38	24	25
				$p^d = 0.03$		
	65+		(n = 71)	(n = 73)	(n = 74)	
		Never	45	68	70	67
		Former	21	16	26	21
		Current	34	15	4	12
				$p^d = 0.05$		

[a]Data for <25 years and for black women not presented because of small sample size.
[b]NC, neighborhood controls; HC, hospital controls; NS, not significant.
[c]See Ref. 6.
[d]p value for chi-square test (2 df) for the difference between the two types of controls.

TABLE 3. Smoking characteristics by sex and case-control status in 775 male and 249 female triplets

Smoking status	Cigarettes per day	Males (n = 775 triplets) Cases (%)	Males NC[a] (%)	Males HC[a] (%)	Females (n = 249 triplets) Cases (%)	Females NC[a] (%)	Females HC[a] (%)
Never		7.5	28.4	32.5	26.1	55.4	60.2
Current		51.6	36.0	25.7	56.6	28.5	19.7
Former		32.5	28.0	32.8	17.3	16.1	20.1
			$p^b = 0.0001$			$p^b = 0.04$	
Ever	1–14	7.6	12.5	10.3	10.8	18.1	13.3
	15–30	41.3	36.4	30.6	41.0	22.1	20.9
	31+	35.2	15.1	17.7	22.1	4.4	5.6
			$p^b = 0.07$			p^b = NS	
Pipe/cigar		8.4	7.6	9.0	0	0	0

[a]NC, neighborhood controls; HC, hospital controls.
[b]p value for chi-square test (2 df) for the difference between the two types of controls.

factors for coronary artery diseases in hospital and neighborhood controls [10]. In males, proportions of current smokers were 12% for hospital patients and 27% for neighbors. Among ever smokers, hospital controls had smoked 30 pack-years versus 33 pack-years in neighborhood controls. In females, proportions of current smokers were 11% for hospital patients and 36% for neighbors and among ever smokers, hospital controls had smoked 22 versus 29 pack-years. Although two studies in addition to ours is a small sample to generalize from, they demonstrate that prevalence of smoking is not necessarily lower in neighborhood controls than in hospital controls, as it is commonly assumed.

The results of the present study may stem from several potential biases, the most serious of which are recall and selection biases. For example, the recall of past exposure in hospital patients, even with non-tobacco-related diseases, may be influenced by having been asked repeatedly whether they smoked and how much. The present study suggests that if there is such recall bias, hospital patients tend to deny their past exposure. Selection bias may have resulted from the lower participation rate of neighborhood controls. Interviewed neighbors may not be typical of all eligible subjects because they were more likely to be home. Indeed, we found that neighbors were more often widowed and had less education on average than hospital controls, that is, they belonged to socioeconomic categories in which current smoking is more prevalent. However, there is no reason to believe that the lower participation rate for neighborhood controls is unique to the present study. This rate is generally ob-

TABLE 4. Conditional logistic regression odds ratios of smoking and several diseases using either neighborhood controls or hospital controls, in males

Disease (no. of triplets)	Smoking status	Cases (n)	NC[a] (n)	HC[a] (n)	Conditional odds ratio[b] NC	Conditional odds ratio[b] HC	ROR (NC:HC)
Lung cancer (n = 248)	Never	6	75	64	1.0	1.0	
	Current	150	81	75	24.2	19.9	1.2
	Former	82	70	91	15.0	8.7	1.7
	Pipes/cigar	10	22	18	6.3	5.3	1.2
Oral cavity cancer (n = 162)	Never	7	36	76	1.0	1.0	
	Current	92	71	28	6.5	38.7	0.2
	Former	39	45	34	4.5	13.5	0.3
	Pipes/cigar	24	10	24	11.7	10.7	1.1
Bladder cancer (n = 147)	Never	15	41	45	1.0	1.0	
	Current	61	55	30	3.0	6.8	0.4
	Former	60	46	58	3.4	3.5	1.0
	Pipes/cigar	11	5	14	5.8	2.1*	2.8
Pancreas cancer (n =42)	Never	7	19	15	1.0	1.0	
	Current	13	10	11	3.6*	3.0*	1.2
	Former	17	7	12	13.1	3.2*	4.1
	Pipes/cigar	5	6	4	1.9*	3.5*	0.5
Myocardial infarction (n = 149)	Never	21	40	41	1.0	1.0	
	Current	74	60	49	2.5	3.4	0.7
	Former	44	38	52	2.3	1.8*	1.3
	Pipes/cigar	10	11	7	1.8*	3.4	0.5

[a]NC, Neighborhood controls; HC, hospital controls; ROR, ratio of odds ratios.
[b]All ORs are statistically significant at $p < 0.05$ except those marked with an asterisk.

TABLE 5. Conditional logistic regression odds ratios of smoking and several diseases using either neighborhood controls or hospital controls, in females

Disease (no. of triplets)	Smoking status	Cases (n)	NC[a] (n)	HC[a] (n)	Odds ratio[b] NC[a]	Odds ratio[b] HC[a]	ROR (NC:HC)
Lung	Never	18	59	61	1.0	1.0	
cancer	Current	76	42	27	7.0	9.4	0.7
(n = 117)	Former	23	16	29	5.0	2.7	1.9
Oral cavity	Never	8	16	21	1.0	1.0	
cancer	Current	17	12	8	3.3	5.4	0.6
(n = 32)	Former	7	4	3	4.4*	5.2	0.8
Bladder	Never	9	6	7	1.0	1.0	
cancer	Current	17	6	4	7.2	16.4	0.4
(n = 31)	Former	5	19	20	1.8*	2.3*	0.8

[a]NC, Neighborhood controls; HC, hospital controls; ROR, ratio of odds ratios.
[b]All ORs are statistically significant at $p < 0.05$ except those marked with an asterisk.

served to be lower than for hospital controls, although it is difficult to compare our rates with those in the literature since, as we showed in Methods and Materials, different definitions of the completion rate could result in estimates ranging from 40 to 76%. It is important to note that, because the association of cigarette smoking and cancers of the lung, oral cavity, and bladder is so strong, the biases related to control selection did not lead to major differences in interpretation of the observed associations. The potential for misinterpretation would have been much more serious if we had been studying weaker associations.

In conclusion, choosing between hospital and neighborhood controls in case-control studies should be dictated by criteria related to the study hypothesis, participation, or cost. In particular, exclusion of hospital controls with diseases known to be tobacco related seems to be a successful strategy for reducing selection bias.

The authors thank Vicki Liang and Emily Chung for data processing, and Marion Moore, who was field supervisor throughout data collection. The study was supported by Grant Nos. CA-17613, CA-32617, and CA-68384 from the National Cancer Institute.

References

1. Hennekens CH, Buring JE. Epidemiology in Medicine. Little, Brown and Co., Boston, 1987, p. 138.
2. Mantel N, Haenszel W. Statistical aspects of the analysis of data from the retrospective studies of disease. J Natl Cancer Inst 1959; 22: 719–748.
3. Wynder EL, Stellman SD. The "over-exposed" control group. Am J Epidemiol 1992;135: 459–461
4. Baron JA, La Vecchia C, Levi F. The antiestrogenic effect of cigarette smoking in women. Am J Obstet Gynecol 1990;162: 502–514.
5. Wynder EL, Stellman SD. The comparative epidemiology of tobacco-related cancers. Cancer Res 1977;37:4608–4622.
6. Center for Disease Control. Cigarette smoking in the United States, 1986. MMWR 1987;36:581–585.
7. SAS/STAT Software: The PHREG Procedure (preliminary documentation). SAS Institute, Inc., Cary, North Carolina, 1991.
8. West DW, Schuman KL, Lyon JL, Robinson LM, Allred R. Differences in risk estimations from a hospital and a population-based case-control study. Int J Epidemiol 1984; 13: 235–239.
9. Stavraky KM, Clarke EA. Hospital or population controls? An unanswered question. J Chron Dis 1983; 36: 301–307.
10. Tell GS, Ryu JE, Thompson CJ, *et al*. Comparison of hospital and neighborhood controls in a study of coronary artery disease. J Clin Epidemiol 1991; 44: 1097–1104.

APPENDIX. Discharge diagnoses of hospital controls, by sex

Diagnosis	Number of diagnoses: Males (n = 775)	Females (n =249)
Stomach cancer	15	2
Colon cancer	70	20
Prostate cancer	82	—
Breast cancer	2	40
Skin cancer	36	18
Cervix cancer	—	12
Leukemias	23	7
Lymphoma	49	11
Sarcoma	22	4
Multiple myeloma	3	2
Endometrial cancer	—	9
Throid cancer	6	1
Ovaries cancer	—	6
Other cancers	28	7
Burns	4	1
Gallbladder + pancreas disease	2	2
Pneumonia	6	1
Urinary tract infection	2	2
Other infectious diseases	38	6
Fractures	25	11
Disk and spine problems	24	7
Other traumatic injuries	24	3
Intestinal hernia	35	8
Arthritis	21	11
Benign prostatic hypertrophy	63	—
Benign neoplastic disease	91	33
Eye disease	19	2
Diabetes	10	1
Other nonneoplastic diseases	55	19
Parkinson's disease	7	—
Appendicitis	6	—
No disease	7	3