

The Use of Interviews with Surrogate Respondents in a Case-Control Study of Oral Cancer

RAYMOND S. GREENBERG, M.D., Ph.D., JONATHAN M. LIFF, Ph.D.,
HELEN R. GREGORY, B.S., A.R.T., AND J. ELAINE BROCKMAN, B.A.

*Departments of Community Health and Biometry, Emory University School of
Medicine, Atlanta, Georgia*

Received December 9, 1985

The purpose of this study was to assess the possible bias that may occur in case-control studies when exposure information is not collected from all potentially eligible cases. The data used in this study were collected in the metropolitan area of Atlanta as part of a multicenter, population-based, case-control study of oropharyngeal cancer. In-person interviews were conducted with 112 cases (67.9 percent) and information on an additional 23 ill or deceased cases (13.9 percent) was collected through surrogate respondents. The cases about whom information was collected from surrogate respondents had more advanced disease at the time of diagnosis and were more likely to be black and less educated than cases who were interviewed in person. Cigarette smoking and consumption of hard liquor were more common among the cases about whom information was collected through surrogates. Therefore, failure to include such information would have resulted in underestimates of the strength of association between these exposures and the risk of oropharyngeal cancer.

Over the past few years, a number of multicenter, case-control studies have been conducted through the resources of population-based cancer registries. In this type of investigation, an attempt is made to identify all residents of particular geographic areas who are diagnosed with the cancer of interest over a specific period of time (viz., cases). Typically, the group of cases is restricted further according to categories of age, race, and sex, as well as other factors, such as the method of diagnostic confirmation. Since the cases are chosen from the general population, the controls are, as well. Area household surveys, random-digit telephone dialing, and lists of residents in areas with community registers are three common methods of population control selection. Because the entire population within a geographic area serves as the sampling frame, the sample is not affected by factors related to the place of hospitalization, such as socioeconomic status and stage of disease [1].

It should be recognized, however, that even in population-based research, the case group may not be representative of all persons with the disease of interest. For example, in the National Bladder Cancer Study [2], a collaborative case-control study conducted in ten geographic areas of the United States, only 73 percent of eligible cases were successfully interviewed. Interview rates were higher for males than for females and for people aged 35-64 years than for older or younger persons. The leading reasons for non-response among cases were death (6.9 percent), disability (7.0 percent), and subject refusal (6.2 percent).

If the non-responding cases differ from the responding cases with regard to exposure

histories, then distorted estimates of effect may be obtained. Depending upon the case selection probabilities involved, and how they differ from control selection probabilities, either an underestimate or an overestimate of the true magnitude of effect might be obtained [3]. One approach to reducing such selection bias is to minimize the amount of missing information through the use of interviews with surrogate respondents for cases who are either deceased or disabled.

The purpose of this study was to compare sociodemographic and exposure characteristics of two groups of cases with oropharyngeal cancer and to measure the potential influence of these differences upon effect estimation in case-control comparisons. The two groups of cases were: (a) those subjects who could be interviewed in person and (b) those subjects for whom information could be obtained only through surrogate respondents.

METHODS

The data for the present analysis were collected in one center of a collaborative, population-based, case-control study of oropharyngeal cancers. This investigation was designed to permit examination of the contribution of various suspected risk factors to the etiology of oropharyngeal malignancies. The source populations for cases and controls for the entire study included: metropolitan Atlanta, Los Angeles, and San Francisco, as well as the state of New Jersey.

In each of these geographic areas, the resources of a population-based cancer registry were used to identify eligible cases. The primary methods of case ascertainment included the review of pathology reports and discharge diagnoses at hospitals in or surrounding the study areas. After a histologically confirmed oral cancer was identified, the medical record was obtained to establish eligibility for participation. The inclusion criteria included age (18 through 79 years), date of first diagnosis (January 1, 1984, through March 31, 1985), residency in one of the study regions, and absence of a prior history of oropharyngeal cancer.

The physician of each patient was contacted to discuss any possible contraindications to contacting the patient (e.g., poor health, emotional distress, mental incompetence). If an interview was considered permissible, then the case was contacted to schedule an appointment.

Controls under 65 years of age were selected from the general population of the study areas using the random-digit telephone dialing technique of Waksberg [4]. Lists provided by the Health Care Financing Administration were used to select controls 65 years of age and older. Controls were selected so that their overall distribution of age (in five-year intervals), sex, race, and geographic area would be similar to that expected for cases. Informed consent was obtained for interviews with all cases and controls.

An attempt was made to conduct an in-person interview with each case and each control. The interviewers in all study areas were uniformly trained and the same questionnaire and interviewing methods were used for cases and controls. Whenever possible, interviews with cases were scheduled within six months of the time of diagnosis in order to minimize recall bias and losses due to death or illness. For cases who were deceased or too ill to participate, an attempt was made to identify an individual who knew the case and was willing to complete the interview. The results reported here are from the metropolitan Atlanta, Georgia, section of this study, which includes Fulton, DeKalb, Gwinnett, Clayton, and Cobb counties, comprising a source population of about 1,700,000 persons in 1980.

In virtually all instances, the interview was conducted at the home of the respondent. In Atlanta, approximately 5 percent of interviews were performed in alternate settings for a variety of reasons, including difficulty in arranging a meeting time at home, recurrent hospitalizations, and concern for the safety of interviewers. For a few interviews, an assistant was present to help overcome communication problems caused by the disease or its treatment.

The interview usually lasted for approximately one hour. Following the interview, the questionnaire was edited for completeness and internal consistency. The reliability of essential information was checked by telephone call back to a 10 percent sample of all respondents. The exposures of interest for the present analysis included a variety of personal behaviors and dietary factors. The key questions in the present context relate to the use of chewing tobacco, multiple vitamins, mouthwash, dentures, beer, hard liquor, and cigarettes. These items were selected because it was felt that surrogate respondents were capable of giving accurate information on these exposures. The specific wording of questions on these items is provided in the Appendix.

The variables were coded to permit categorical data analysis: age (<65 years versus ≥ 65 years), race (white versus black), sex (male versus female), county of residence (Fulton versus all others), marital status (married versus all others), education (≤ 12 years versus > 12 years), stage of disease (localized versus all others), and each of the aforementioned exposures (yes versus no). The odds ratio [5] was used to measure the strength of association between independent variables and type of interview (personal versus surrogate) among cases. Corresponding approximate 95 percent test-based confidence intervals around the odds ratios [6] also were calculated.

To evaluate the influence of the detected sociodemographic and exposure differences on effect estimation, two sets of case-control comparisons were made. First, only cases who were interviewed in person were compared to controls. Then, all cases (including those for whom information was collected from surrogate respondents) were compared to controls. Analyses were stratified separately by age, race, sex, county of residence, marital status, educational level, and stage of disease. Summary odds ratios were estimated with the method of Mantel and Haenszel [7] and corresponding approximate 95 percent test-based confidence intervals [6] were calculated. The presence of confounding was determined by meaningful differences in the crude and adjusted odds ratios [8].

RESULTS

In Atlanta, a total of 165 eligible subjects were identified for this investigation (Table 1). Of the eligible cases, 112 (67.9 percent) were interviewed in person, 23 (13.9 percent) could not be interviewed in person, so interviews were conducted with surrogate respondents, and 30 cases (18.1 percent) had no interviews conducted. The reasons for non-response included: death (6.0 percent), illness (6.0 percent), refusal (2.4 percent), and inaccessibility (3.6 percent).

A summary of the types of surrogate respondents included in this study is provided in Table 2. In each case, effort was made to identify the individual with the greatest knowledge about the behaviors of the relevant subject. Therefore, the highest priority in the selection of surrogates was assigned to persons who lived with the subject during the case's adult life. Almost two-thirds of the surrogates were either spouses/primary intimates, or offspring of the cases. The remainder of interviews with surrogate respondents were completed with parents, siblings, more remote relatives, or acquaintances of the cases.

TABLE 1
Summary of Cases from Atlanta

	<i>n</i>	(%)
Personal interview	112	(67.9)
Deceased	31	(18.8)
Surrogate interview	21	(12.7)
No interview	10	(6.0)
Critically ill	12	(7.3)
Surrogate interview	2	(1.2)
No interview	10	(6.0)
Interview refused	4	(2.4)
Unavailable or moved	6	(3.6)
Total	165	(100)

As indicated in Table 3, the principal difference between cases interviewed in person and those who had interviews with surrogates was the stage of their disease at diagnosis. Almost 40 percent of cases interviewed in person had localized disease, compared with only 4 percent of cases who had surrogate interviews. This finding is not surprising, since the indications for conducting a surrogate interview included illness or death of the case, and stage of disease is a strong determinant of morbidity and mortality. The cases who were interviewed in person also included a significantly higher proportion of whites than did the cases who had interviews with surrogates. Lower educational levels and residence within Fulton County tended to be more common among cases with surrogate interviews.

The exposure histories of the two types of cases are contrasted in Table 4. Although none of these differences reached statistical significance, a greater proportion of the cases for whom surrogate interviews were required consumed hard liquor and used cigarettes than did cases who were interviewed in person. To evaluate the impact of these exposure differences on effect estimation, case-control comparisons were performed with and without the inclusion of information from surrogate respondents.

The crude odds ratios and corresponding 95 percent confidence intervals for case-control analyses are presented in Table 5. The inclusion of information from surrogate respondents had little impact on the estimated strengths of association for all exposures except hard liquor and cigarette smoking. For the latter two exposures, inclusion of information from surrogates resulted in effect estimates that were appreciably greater than the estimates obtained without surrogates.

To evaluate the possible role of confounding by age, race, sex, educational level, and

TABLE 2
Summary of Types of Surrogate Respondents for Oral Cancer

	<i>n</i>	(%)
Spouse/primary intimate	9	(39.1)
Offspring	6	(26.1)
Parent/sibling	5	(21.7)
Distant relative	2	(8.7)
Acquaintance	1	(4.3)
Total	23	(100)

TABLE 3
Comparisons of Characteristics of Oral Cancer Cases by Type of Interview

Characteristic	Personal Interview		Surrogate Interview		OR	95% CI
	<i>n</i>	(%)	<i>n</i>	(%)		
Total	112	(100.0)	23	(100.0)		
Age (<65)	73	(65.2)	13	(56.5)	1.4	(0.6, 3.6)
Race (White)	78	(69.6)	11	(47.8)	2.5	(1.0, 6.1)
Sex (Male)	73	(65.2)	16	(69.6)	0.8	(0.3, 2.2)
County (Fulton)	62	(55.4)	17	(73.9)	0.4	(0.2, 1.2)
Marital Status (Married)	68	(60.7)	11	(47.8)	1.7	(0.7, 4.1)
Education (<12)	45	(40.2)	14	(60.9)	0.4	(0.2, 1.1)
Stage (Local)	43	(38.4)	1	(4.3)	13.9	(2.8, 70.3)

marital status, stratified analyses were performed for each of the associations depicted in Table 5. Confounding was detected in only one circumstance: the effect of chewing tobacco was confounded by educational level. The crude and education-adjusted odds ratios for chewing tobacco are presented in Table 6. The small sample size did not allow simultaneous adjustment of several factors.

The reliability of information on cigarette smoking collected from surrogate respondents was assessed by comparing information from interviews with smoking behavior documented in medical records. There was agreement between the smoking status reported from these two sources in 22 of 23 instances (95.6 percent). An evaluation of the reliability of alcohol consumption information reported by surrogate respondents could not be performed since this information was not routinely reported in medical records.

DISCUSSION

As demonstrated in this investigation, oral cancer cases for whom personal interviews could be conducted tended to have less advanced disease than cases for whom information was collected through surrogate respondents. The cases who were interviewed in person also tended to be of higher socioeconomic status, as indicated by educational level (above 12 years), and differed by race (more were white), and place

TABLE 4
Comparisons of Exposure Histories of Oral Cancer Cases by Type of Interview

Exposure	Personal Interview		Surrogate Interview		OR	95% CI
	<i>n</i>	(%)	<i>n</i>	(%)		
Total	112	(100.0)	23	(100.0)		
Mouthwash	70	(62.5)	11	(47.8)	1.7	(0.7, 4.2)
Chewing tobacco	9	(8.0)	0	(0)		
Multiple vitamins	42	(37.5)	6	(26.1)	1.7	(0.6, 4.6)
Dentures	57	(50.9)	10	(43.5)	1.3	(0.5, 3.3)
Beer	80	(71.4)	16	(69.6)	1.1	(0.4, 2.9)
Hard liquor	91	(81.3)	22	(95.7)	0.2	(0.0, 1.3)
Cigarettes	90	(80.4)	22	(95.7)	0.2	(0.0, 1.2)

TABLE 5
Crude Oral Cancer Case-Control Comparisons for Exposures, with Information from Surrogate Interviews Excluded and Included

Exposure	Surrogates Excluded		Surrogates Included	
	OR	95% CI	OR	95% CI
Chewing tobacco	0.8	(0.3, 1.8)	0.6	(0.3, 1.5)
Multiple vitamins	1.0	(0.6, 1.6)	0.9	(0.6, 1.5)
Mouthwash	1.2	(0.7, 2.0)	1.1	(0.7, 1.8)
Dentures	1.4	(0.8, 2.3)	1.3	(0.8, 2.1)
Beer	1.8	(1.1, 3.1)	1.8	(1.1, 3.0)
Hard liquor	2.2	(1.2, 3.9)	2.6	(1.5, 4.6)
Cigarettes	3.2	(1.8, 5.6)	3.8	(2.2, 6.6)

of residence (more were likely to live outside of Fulton County). Elevated rates of morbidity and mortality among socially disadvantaged persons have been observed for individuals with oral cancer [9], as well as patients with other types of cancer [9,10]. Lack of access to medical care, lack of knowledge or motivation to seek such care, and poor nutritional status are three possible explanations for poorer survival. The lower socioeconomic status of cases requiring surrogate interviews is, therefore, no surprise.

When specific exposures were considered, higher rates of cigarette smoking and alcohol consumption were found among cases who were not interviewed in person. There was only one non-smoker and one non-consumer of hard liquor among the 23 cases for whom surrogate interviews were required. This finding is consistent with the observed elevated prevalences of cigarette smoking [11] and heavy alcohol consumption [12] among the socially disadvantaged. It has been demonstrated that persons who consume moderate to large amounts of alcohol also tend to be smokers [13].

Because low socioeconomic status is a risk factor associated with case unavailability for interview, and is also associated with the primary risk factors for the disease of interest, alcohol consumption and cigarette smoking, the exclusion of cases who were not interviewed in person resulted in an underestimation of the measures of association. In the present study this response bias reduced the odds ratio from 3.8 to 3.2 for cigarette smoking, and from 2.6 to 2.2 for hard liquor consumption. Odds ratios for exposures less strongly related to availability for interview or to disease were affected to a smaller degree by the exclusion of surrogate interviews, although the statistical precision of such estimates was reduced.

There are two possible explanations for apparent differences between cases requiring surrogate interviews and those who can be interviewed in person. The discussion

TABLE 6
Crude and Education-Adjusted Associations Between Chewing Tobacco and Oral Cancer, With and Without the Inclusion of Information from Surrogate Interviews

Adjustment Variable	Surrogates Excluded		Surrogates Included	
	OR	95% CI	OR	95% CI
None	0.8	(0.3, 1.8)	0.6	(0.3, 1.5)
Education	0.5	(0.2, 1.2)	0.4	(0.1, 0.9)

above assumed that the differences were real and were related to prognostic factors for the disease. The other possibility is that the perceived differences in exposure prevalence can be explained by differences in reporting between surrogates and cases. In the present analysis, validation of surrogate responses about cigarette smoking was performed with an independent source of information. Since no validation was performed on the responses of cases, it is possible that apparent discrepancies in exposure frequency were attributable to systematic underreporting of certain exposures among cases.

Clearly, information from surrogates about some exposures, e.g., cigarette consumption or reproductive history, may be better than that about others, e.g., childhood illnesses. To the extent that an investigator can validate surrogate information, either by utilizing additional sources of information or by obtaining both in-person and surrogate information from a sample of subjects, a prudent decision about whether to include or exclude surrogate information can be reached.

The results presented here represent only the Atlanta metropolitan area portion of a population-based collaborative case-control study involving oropharyngeal cancer in four geographic areas. The observed findings require confirmation with the full data set from this study, and further substantiation in other epidemiologic investigations.

APPENDIX

Questions Used to Collect Information on the Exposures Under Investigation

1. Before one year ago, did (you/your——) ever chew tobacco for six months or more?
 2. During (you/your——'s) adult life, before one year ago, did (you/he/she) ever take a multiple vitamin supplement such as One-A-Day pills on a regular basis for six months or longer?
 3. Before one year ago, did (you/your——) ever use mouthwash on a regular basis? By regular basis we mean at least once a week for six months or more.
 4. Before one year ago, did (you/he/she) have removable full or partial dentures for (your/his/her) upper jaw?
 5. Before one year ago, did (you/he/she) ever have a total of 20 drinks of beer over (your/his/her) entire life?
 6. Before one year ago, did (you/he/she) ever have a total of 20 drinks of hard liquor, brandy, or liqueurs over (your/his/her) entire life?
 7. Before one year ago, did (you/your——) ever smoke a total of 100 or more cigarettes?
-

ACKNOWLEDGEMENTS

The authors wish to thank the physicians and hospitals who provided access to subjects involved. This work was supported in part by contracts NO1-CN-55429 and NO1-CP-31041-01 from the National Cancer Institute.

Ms. Florence Wilson and Ms. Lisa Shalet assisted in data collection, and Ms. Gail Smith entered the data into a computer database. Ms. Pauline Storey assisted in the preparation of the manuscript.

REFERENCES

1. Sackett DL: Bias in analytic research. *J Chron Dis* 32:51-63, 1979
2. Hartge P, Cahill JI, West D, et al: Design and methods in a multi-center case-control interview study. *Am J Public Health* 74:52-56, 1984

3. Kleinbaum DG, Morgenstern H, Kupper LL: Selection bias in epidemiologic studies. *Am J Epidemiol* 113:452–463, 1981
4. Waksberg J: Sampling methods for random digit dialing. *J Am Stat Assoc* 73:40–46, 1978
5. Cornfield J: A statistical problem arising from retrospective studies. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*. Vol 4. Edited by J Neyman. Berkeley, CA, University of California Press, 1956, pp 135–148
6. Miettinen OS: Estimability and estimation in case-referent studies. *Am J Epidemiol* 103:226–235, 1976
7. Mantel N, Haenszel W: Statistical aspects of the analysis of data from retrospective studies of disease. *J NCI* 22:719–748, 1959
8. Kleinbaum DG, Kupper LL, Morgenstern H: *Epidemiologic Research: Principles and Quantitative Methods*. Belmont, CA, Lifetime Learning Publications, 1982, pp 242–265
9. Berg JW, Ross R, Latourette HB: Economic status and survival of cancer patients. *Cancer* 39:467–477, 1977
10. Chirikos TN, Reiches NA, Moeschberger ML: Economic differentials in cancer survival: a multivariate analysis. *J Chron Dis* 37:183–193, 1984
11. Borland BR, Rudolph JP: Relative effects of low socioeconomic status, parent smoking, and poor scholastic performance on smoking among high school students. *Soc Sci Med* 9:27–30, 1975
12. Kozarevic D, Racic Z, Gordon T, et al: Drinking habits and other characteristics. *Am J Epidemiol* 116:287–301, 1982
13. Cahalam D, Cisin IH, Crossley HM: *American drinking practices: National study of drinking behavior and attitudes*. New Brunswick, NJ, Rutgers Center of Alcohol Studies, 1969