Centers for Disease Control and Prevention
Epidemiology Program Office
Case Studies in Applied Epidemiology
No. 731-703

Cigarette Smoking and Lung Cancer

Student's Guide

Learning Objectives

After completing this case study, the participant should be able to:

- ☐ Discuss the elements of study design, and the advantages and disadvantages of case-control versus prospective cohort studies;
- ☐ Discuss some of the biases that might have affected these studies;
- ☐ Calculate a rate ratio, rate difference, odds ratio, and attributable risk percent;
- ☐ Interpret each measure and describe each measure's main use; and
- ☐ Review the criteria for causation.

This case study is based on the classic studies by Doll and Hill that demonstrated a relationship between smoking and lung cancer. Two case studies were developed by Clark Heath, Godfrey Oakley, David Erickson, and Howard Ory in 1973. The two case studies were combined into one and substantially revised and updated by Nancy Binkin and Richard Dicker in 1990. Current version updated by Richard Dicker with input from Julie Magri and the 2003 EIS Summer Course instructors.

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Public Health Service

A causal relationship between cigarette smoking and lung cancer was first suspected in the 1920s on the basis of clinical observations. To test this apparent association, numerous epidemiologic studies were undertaken between 1930 and 1960. Two studies were conducted by Richard Doll and Austin Bradford Hill in Great Britain. The first was a case-control study begun in 1947 comparing the smoking habits of lung cancer patients with the smoking habits of other patients. The second was a cohort study begun in 1951 recording causes of death among British physicians in relation to smoking habits. This case study deals first with the case-control study, then with the cohort study.

Data for the case-control study were obtained from hospitalized patients in London and vicinity over a 4-year period (April 1948 - February 1952). Initially, 20 hospitals, and later more, were asked to notify the investigators of all patients admitted with a new diagnosis of lung cancer. These patients were then interviewed concerning smoking habits, as were controls selected from patients with other disorders (primarily non-malignant) who were hospitalized in the same hospitals at the same time.

Data for the cohort study were obtained from the population of all physicians listed in the *British Medical Register* who resided in England and Wales as of October 1951. Information about present and past smoking habits was obtained by questionnaire. Information about lung cancer came from death certificates and other mortality data recorded during ensuing years.

Question 1: What makes the first study a case-control study?

Question 2: What makes the second study a cohort study?

The remainder of Part I deals with the case-control study.

Question 3: Why might hospitals have been chosen as the setting for this study?

Question 4: What other sources of cases and controls might have been used?

Question 5: What are the advantages of selecting controls from the same hospitals as cases?

Question 6: How representative of all persons with lung cancer are hospitalized patients with lung cancer?

Question 7: How representative of the general population without lung cancer are hospitalized patients without lung cancer?

Question 8: How may these representativeness issues affect interpretation of the study's results?

Over 1,700 patients with lung cancer, all under age 75, were eligible for the case-control study. About 15% of these persons were not interviewed because of death, discharge, severity of illness, or inability to speak English. An additional group of patients were interviewed but later excluded when initial lung cancer diagnosis proved mistaken. The final study group included 1,465 cases (1,357 males and 108 females).

The following table shows the relationship between cigarette smoking and lung cancer among male cases and controls.

Table 1. Smoking status before onset of the present illness, lung cancer cases and matched controls with other diseases, Great Britain, 1948-1952.

	Cases	Controls
Cigarette smoker	1,350	1,296
Non-smoker	7	61
Total	1,357	1,357

Question 9: From this table, calculate the proportion of cases and controls who smoked.

Proportion smoked, cases:

Proportion smoked, controls:

Question 10: What do you infer from these proportions?

Question 11a: Calculate the odds of smoking among the cases.

Question 11b: Calculate the odds of smoking among the controls.

Question 12: Calculate the ratio of these odds. How does this compare with the cross-product ratio?

Question 13: What do you infer from the odds ratio about the relationship between smoking and lung cancer?

Table 2 shows the frequency distribution of male cases and controls by average number of cigarettes smoked per day.

Table 2. Most recent amount of cigarettes smoked daily before onset of the present illness, lung cancer cases and matched controls with other diseases, Great Britain, 1948-1952.

Daily number of cigarettes	# Cases	# Controls	Odds Ratio
0	7	61	referent
1-14	565	706	___
15-24	445	408	___
25+	340	182	___
All smokers	1,350	1,296	___
Total	1,357	1,357	

Question 14: Compute the odds ratio by category of daily cigarette consumption, comparing each smoking category to nonsmokers.

Question 15: Interpret these results.

Although the study demonstrates a clear association between smoking and lung cancer, cause-and-effect is not the only explanation.

Question 16: What are the other possible explanations for the apparent association?

The next section of this case study deals with the cohort study.

Data for the cohort study were obtained from the population of all physicians listed in the British Medical Register who resided in England and Wales as of October 1951. Questionnaires were mailed in October 1951, to 59,600 physicians. The questionnaire asked the physicians to classify themselves into one of three categories: 1) current smoker, 2) ex-smoker, or 3) nonsmoker. Smokers and ex-smokers were asked the amount they smoked, their method of smoking, the age they started to smoke, and, if they had stopped smoking, how long it had been since they last smoked. Nonsmokers were defined as persons who had never consistently smoked as much as one cigarette a day for as long as one year.

Usable responses to the questionnaire were received from 40,637 (68%) physicians, of whom 34,445 were males and 6,192 were females.

Question 17: How might the response rate of 68% affect the study's results?

The next section of this case study is limited to the analysis of male physician respondents, 35 years of age or older.

The occurrence of lung cancer in physicians responding to the questionnaire was documented over a 10-year period (November 1951 through October 1961) from death certificates filed with the Registrar General of the United Kingdom and from lists of physician deaths provided by the British Medical Association. All certificates indicating that the decedent was a physician were abstracted. For each death attributed to lung cancer, medical records were reviewed to confirm the diagnosis.

Diagnoses of lung cancer were based on the best evidence available; about 70% were from biopsy, autopsy, or sputum cytology (combined with bronchoscopy or X-ray evidence); 29% were from cytology, bronchoscopy, or X-ray alone; and only 1% were from just case history, physical examination, or death certificate.

Of 4,597 deaths in the cohort over the 10-year period, 157 were reported to have been caused by lung cancer; in 4 of the 157 cases this diagnosis could not be documented, leaving 153 confirmed deaths from lung cancer.

The following table shows numbers of lung cancer deaths by daily number of cigarettes smoked at the time of the 1951 questionnaire (for male physicians who were nonsmokers and current smokers only). Person-years of observation ("person-years at risk") are given for each smoking category. The number of cigarettes smoked was available for 136 of the persons who died from lung cancer.

Table 3. Number and rate (per 1,000 person-years) of lung cancer deaths by number of cigarettes smoked per day, Doll and Hill physician cohort study, Great Britain, 1951-1961.

Daily number of cigarettes smoked	Deaths from lung cancer	Person-years at risk	Mortality rate per 1000 person-years	Rate Ratio	Rate difference per 1000 person-years
0	3	42,800	0.07	referent	referent
1-14	22	38,600	____	____	____
15-24	54	38,900	____	____	____
25+	57	25,100	____	____	____
All smokers	133	102,600	____	____	____
Total	136	145,400	____		

Question 18: Compute lung cancer mortality rates, rate ratios, and rate differences for each smoking category. What do each of these measures mean?

Question 19: What proportion of lung cancer deaths among all smokers can be attributed to smoking? What is this proportion called?

Question 20: If no one had smoked, how many deaths from lung cancer would have been averted?

The cohort study also provided mortality rates for cardiovascular disease among smokers and nonsmokers. The following table presents lung cancer mortality data and comparable cardiovascular disease mortality data.

Table 4. Mortality rates (per 1,000 person-years), rate ratios, and excess deaths from lung cancer and cardiovascular disease by smoking status, Doll and Hill physician cohort study, Great Britain, 1951-1961.

	Mortality rate per 1,000 person-years					
	Smokers	Non-smokers	All	Rate ratio	Excess deaths per 1,000 person-years	Attributable risk percent among smokers
Lung cancer	1.30	0.07	0.94	18.5	1.23	95%
Cardiovascular disease	9.51	7.32	8.87	1.3	2.19	23%

Question 21: Which cause of death has a stronger association with smoking? Why?

In calculating the **attributable risk percent**, the excess lung cancer deaths attributable to smoking is expressed as a percentage of all lung cancer mortality among all smokers. The attributable risk percent of 95% for smoking may be interpreted as the proportion of lung cancer deaths among smokers that could have been prevented if they had not smoked.

A similar measure, the **population attributable risk percent** expresses the excess lung cancer deaths attributable to smoking as a percentage of all lung cancer mortality among the entire population. From a prevention perspective, the population attributable risk percent for a given exposure can be interpreted as the proportion of cases in the entire population that would be prevented if the exposure had not occurred. The population attributable risk percent is often used in assessing the cost-effectiveness and cost-benefit of community-based intervention programs.

One formula for the population attributable risk percent is:

PAR% = (Incidence in entire population – Incidence in unexposed) / Incidence in entire population

Question 22: Calculate the population attributable risk percent for lung cancer mortality and for cardiovascular disease mortality. How do they compare? How do they differ from the attributable risk percent?

Question 23: How many lung cancer deaths per 1,000 persons per year are attributable to smoking among the entire population? How many cardiovascular disease deaths?

The following table shows the relationship between smoking and lung cancer mortality in terms of the effects of stopping smoking.

Table 5. Number and rate (per 1,000 person-years) of lung cancer deaths for current smokers and ex-smokers by years since quitting, Doll and Hill physician cohort study, Great Britain, 1951-1961.

Cigarette smoking status	Lung cancer deaths	Rate per 1000 person-years	Rate Ratio
Current smokers	133	1.30	18.5
For ex-smokers, years since quitting:			
<5 years	5	0.67	9.6
5-9 years	7	0.49	7.0
10-19 years	3	0.18	2.6
20+ years	2	0.19	2.7
Nonsmokers	3	0.07	1.0 (ref)

Question 24: What do these data imply for the practice of public health and preventive medicine?

As noted at the beginning of this case study, Doll and Hill began their case-control study in 1947. They began their cohort study in 1951. The odds ratios and rate ratios from the two studies by numbers of cigarettes smoked are given in the table below.

Table 6. Comparison of measures of association from Doll and Hill's 1948-1952 case-control study and Doll and Hill's 1951-1961 physician cohort study, by number of cigarettes smoked daily, Great Britain.

Daily number of Cigarettes smoked	Rate ratio from cohort study	Odds ratio from case-control study
0	1.0 (ref)	1.0 (ref)
1-14	8.1	7.0
15-24	19.8	9.5
25+	32.4	16.3
All smokers	18.5	9.1

Question 25: Compare the results of the two studies. Comment on the similarities and differences in the computed measures of association.

Question 26: What are the advantages and disadvantages of case-control vs. cohort studies?

Answer 26

	Case-control	Cohort
Sample size		
Costs		
Study time		
Rare disease		
Rare exposure		
Multiple exposures		
Multiple outcomes		
Progression, spectrum of illness		
Disease rates		
Recall bias		
Loss to follow-up		
Selection bias		

Question 27: Which type of study (cohort or case-control) would you have done first? Why? Why do a second study? Why do the other type of study?

Question 28: Which of the following criteria for causality are met by the evidence presented from these two studies?

Answer 28

	YES	NO
Strong association		
Consistency among studies		
Exposure precedes disease		
Dose-response effect		
Biologic plausibility		

REFERENCES

1. Doll R, Hill AB. Smoking and carcinoma of the lung. *Brit Med J* 1950; 2:739-748.

2. Doll R, Hill AB. A study of the aetiology of carcinoma of the lung. *Brit Med J* 1952; 2:1271-1286.

3. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits. *Brit Med J* 1954; 1:1451-1455.

4. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking. *Brit Med J* 1956; 2:1071-1081.

5. Doll R, Hill AB. Mortality in relation to smoking: 10 years' observation of British doctors. *Brit Med J* 1964; 1:1399-1410, 1460-1467.

6. U. S. Public Health Service. Smoking and health. Report of the Advisory Committee to the Surgeon General of the Public Health Service. US Department of Health, Education, and Welfare, PHS, CDC. PHS Publication No. 1103, 1964.

7. Hill AB. The environment and disease: association or causation? *Proc R Soc Med* 1965;58:295-300.

8. Levy RA, Marimont RB. Lies, damned lies, and 400,000 smoking-related deaths. *Regulation* 1998; 21-29.