Cigarette Smoking and Lung Cancer*

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 1

(Time to completion: 10 minutes)

In the 1920s, health care workers in Great Britain first began to suspect a relationship between cigarette smoking and lung cancer. The suspicion was based on the fact that many patients who acquired lung cancer were also smokers. Although this was an astute observation, these workers lacked the scientific evidence to justify their position. As a result, between 1930 and 1960, numerous epidemiologic studies were undertaken to try to quantify the relationship between cigarette smoking and lung cancer.

Two of these studies, one in 1947 by Sir Richard Doll and one in 1951 by A.B. Hill, are considered classics. Doll used the case-control study method and compared the smoking history of a group of hospitalized patients with lung cancer with the smoking history of a similar group without lung cancer. Hill used a cohort study, categorizing a group of British physicians according to their smoking histories and then analyzing the causes of death among those who died, to see whether cigarette smokers had the highest incidence of lung cancer.

You are an EIS (Epidemic Intelligence Service) Officer, or CDC Disease Detective, whose mission it is to analyze these studies in order to assess the relationship between cigarette smoking and lung cancer.

○ **1. What is the purpose of an epidemiologic study?**

*This exercise is adapted from a case study used each year in CDC's Epidemic Intelligence Service (EIS) Summer Course, which trains incoming EIS Officers. This adaptation was created by Dr. Natale A. Carasali while at CDC in spring 2001, fulfilling a medical epidemiology elective during his final year of training at the University of Nevada

School of Medicine.

Cigarette Smoking and Lung Cancer

○ **2. What is the hypothesis both Doll and Hill are trying to test?**

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 2

(Time to completion: 15 minutes)

Doll’s case-control study: Data for this study were collected from hospitalized patients in London and the surrounding communities over a four-year period (April 1948 to February 1952). Hospital personnel at more than 20 hospitals were asked to contact investigators whenever a patient was admitted with a new diagnosis of lung cancer. These patients were then interviewed about their smoking habits. At the same time, investigators interviewed a random sample of patients from the same hospitals, but with different illnesses, about their smoking habits.

Hill’s cohort Study: Data for this study were obtained from physicians listed in the British Medical Register who resided in England and Wales as of October 1951. At the beginning of the study, a questionnaire was used to collect information about the physicians’ past and present smoking habits. They were then categorized according to their exposure to cigarette smoke. In the ensuing years, investigators gathered information about deaths attributed to lung cancer from death certificates and other mortality data.

○ **3. In the first study, why were investigators interested in interviewing patients who were hospitalized for disorders other than lung cancer?**

○ **4. What makes the second study a cohort study?**

○ **5. Suppose you are one of Doll's investigators, and you get a call from a local hospital informing you of a newly diagnosed case of lung cancer. What is the first thing you should do?** (Tip: Don't forget your basic epidemiology.)

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 3
(Parts 3-6 focus on Doll's case-control study)
(Time to completion: 15 minutes)

One of the most important decisions you make in a case-control study is deciding who the controls should be. Ideally, case-subjects and controls should differ only with respect to disease status and exposure to the agent under investigation. This degree of similarity is rarely possible, but investigators make every attempt to match the case-subjects and controls as closely as possible.

○ 6. Why is it important to try to match case-subjects and controls so closely?

○ 7. What are the advantages of selecting controls from the same hospitals as case-subjects?

Cigarette Smoking and Lung Cancer

○ **8. Why do you think Doll and his colleagues conducted their study in hospitals?**

○ **9. Could investigators have chosen case-subjects and controls from sources besides hospitals? If so, please list some examples.**

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 4
(Parts 3-6 focus on Doll's case-control study)
(Time to completion: 15 minutes)

In epidemiologic studies, it is important that the case-subjects and controls closely match the populations they are intended to represent. When Doll conducted his case-control study, the great majority of lung cancer patients required hospitalization. Therefore, the case-subjects in the study are very representative of the region's lung cancer population between 1948 and 1952.

○ **10. What population are the controls supposed to represent? How similar are they to this population?**

○ **11. What effect will the controls have on the study's results, given the quality of their match with the population they represent?**

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 5

(Parts 3-6 focus on Doll's case-control study)

(Time to completion: 45 minutes)

Over 1700 people with lung cancer, all under the age of 75, were eligible for the case-control study. Approximately 15% were not interviewed because of death, severity of illness, discharge from the hospital, or inability to speak English. An additional group of patients was interviewed for the study but was later excluded when their initial diagnosis of lung cancer proved to be wrong. The final study group included 1465 case-subjects (1357 men and 108 women). Only men were included in the study.

Table 1

	Cases	**Controls**
Cigarette Smokers	1,350	1,296
Nonsmokers	7	61
Total	1,357	1,357

○ **12. From Table 1, calculate the proportion of case-subjects and controls who smoked.**

○ **13. How would you interpret these proportions?**

○ **14. Calculate the odds of smoking for both case-subjects and controls.**

○ **15. Calculate the odds ratio.**

○ **16. What does this odds ratio tell us about the relationship between smoking and lung cancer?**

PART 6

(Parts 3-6 focus on Doll's case-control study)

(Time to completion: 30 minutes)

The following table shows the frequency distribution of case-subjects and controls by the average number of cigarettes smoked per day.

Table 2

Cigarettes smoked daily	Cases	Controls	Odds Ratio
0	7	61	Referent
1-14	565	706	
5-24	445	408	
25 +	340	182	
All smokers	1,350	1,296	
Total	1,357	1,357	

○ 17. Complete Table 2 by calculating the odds ratio for each category of cigarette consumption.

Cigarette Smoking and Lung Cancer

○ **18. How would you interpret these results?**

○ **19. While the study appears to demonstrate a clear association between smoking and lung cancer, cause and effect is not the only possible explanation. List other possibilities.**

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 7
(Parts 7–11 focus on Hill's cohort study)
(Time to completion: 10 minutes)

You may recall that 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.

In October 1951, questionnaires were mailed to 59,600 physicians. The questionnaire asked the physician to classify him/herself into one of three categories: 1) current smoker; 2) ex-smoker; or 3) nonsmoker. Smokers and ex-smokers were asked how much they smoked, their method of smoking, the age at which they started to smoke, and if they had stopped smoking, how long since they last smoked. Nonsmokers were defined as people who had never consistently smoked as much as one cigarette per day for as long as one year.

Usable responses were received from 40,637 (68%) of the physicians (34,445 men and 6,192 women).

○ **20. How do you think the 68% response rate might affect the study's results?**

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 8
(Parts 7–11 focus on Hill's cohort study)
(Time to completion: 20 minutes)

For this section, we will focus only on male physician respondents 35 years of age or older.

The occurrence of lung cancer among respondents was documented from November 1951 through October 1961. Over this ten-year period, 4,957 people in the cohort died, 157 from lung cancer. Four of these 157 deaths could not be documented, leaving 153 confirmed deaths from lung cancer.

The following table provides the number of deaths from lung cancer by number of cigarettes smoked per day. (The daily smoking rate was available for only 136 of the 153 decedents). Person-years at risk are given for each smoking category.

Table 3

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			

*See note on next page for definitions.

Note:

Person-Years at Risk: The sum of the amount of time that each person in a study was observed. If a person drops out of a study or is lost to follow-up for some reason, only the time that person was involved is counted.

Mortality Rate: A measure of the frequency with which death occurs in a defined population during a specified period of time.

Rate Ratio: The mathematical comparison of two groups by incidence rates, person-time rates, or mortality rates. The rate for the group of primary interest is divided by the rate for the comparison group.

Rate Difference: The difference in the mortality rates of two groups. It indicates the excess risk of developing a disease that is attributable to the exposure.

○ 21. Complete Table 3 by calculating lung cancer mortality rates, rate ratios and rate differences for each smoking category. What is the overall trend concerning lung cancer mortality rates? How would you interpret your findings for the rate ratio and rate difference categories?

Note:

The **attributable risk percent (AR%)**, also know as the attributable proportion, is the proportion of disease in a group that can be attributed to a particular exposure. Attributable risk percent can be calculated two ways:

$$AR\% = (Incidence_{exposed} – Incidence_{unexposed}) / Incidence_{exposed} \times 100\%$$

or

$$AR\% = (Rate\ Ratio – 1) / Rate\ Ratio \times 100\%$$

O **22. Calculate the attributable risk percent for this study. What does the answer tell us?**

○ **23. Given your answer to question 21, how many deaths from lung cancer could have been avoided if no one had smoked?**

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 9
(Parts 7–11 focus on Hill's cohort study)
(Time to completion: 30 minutes)

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

Table 4

	Mortality Rate per 1000 Person-Years			Rate Ratio for All Smokers	Excess Deaths per 1000 Person-Years	Attributable Risk Percent
	All Smokers	**Zero Cigarettes Daily**	**Total**			
Lung Cancer	1.30	0.07	0.94	18.6	1.23	95%
Cardiovascular Disease	9.51	7.32	8.87	1.3	2.19	23%

○ 24. Given the data in Table 4, with which cause of death is smoking more strongly associated? Why?

○ 25. From Table 4, calculate the population attributable risk percent for lung cancer mortality and for cardiovascular disease mortality. What do these values tell us?

Note:

Population attributable risk percent (PAR%) measures excess death caused by a particular risk factor. By "excess death," we mean the number (incidence) of deaths from a certain risk factor in a population that is over and above the number in the unexposed group. PAR% is calculated several ways, including:

$$\textbf{PAR\%} = \frac{\text{Incidence}_{\text{entire study pop.}} - \text{Incidence}_{\text{unexposed pop.}}}{\text{Incidence}_{\text{entire study pop.}}} \times 100\%$$

PAR% = (% exposed among cases) x (Rate Ratio -1) ÷ Rate Ratio

○ **26. In the entire study population, how many deaths due to lung cancer per 1,000 people per year (1,000 person-years) can be attributed to smoking? How many deaths from cardiovascular disease?**

Cigarette Smoking and Lung Cancer

A *Disease Detectives* Exercise from the
Centers for Disease Control and Prevention

PART 10
(Parts 7–11 focus on Hill's cohort study)
(Time to completion: 10 minutes)

The following table presents deaths due to lung cancer according to the decedents' cigarette smoking status.

Table 5

Cigarette Smoking Status	Number of Lung Cancer Deaths	Mortality Rate per 1,000 Person-Years	Rate Ratio
Current smokers	133	1.30	18.6
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.)

○ 27. What do the data in Table 5 suggest regarding smokers, nonsmokers, and ex-smokers? What does this imply from a public health perspective?

PART 11
(Parts 7–11 focus on Hill's cohort study)
(Time to completion: 25 minutes)

The following table presents results from the two studies, by the number of cigarettes respondents smoked per day.

Table 6

Number of Cigarettes Smoked Daily	Odds Ratio from Case-Control Study	Rate Ratio from Cohort Study
0	1.0 (ref.)	1.0 (ref.)
1-14	7.0	8.1
15-24	9.5	19.8
25+	16.3	32.4
All smokers	18.5	9.1

O **28. What is similar about the results of Doll's case-control study and Hill's cohort study? What is different? How would you account for this difference?**

○ **29. Compare case-control studies and cohort studies in terms of the following requirements:**

	Case-Control	Cohort
Sample size		
Cost		
Study time		

○ **30. Which type of study would be an advantage and which would be a disadvantage in investigations intended to study the following issues:**

	Case-Control	Cohort
Rare disease		
Rare exposure		
Multiple exposures		
Multiple outcomes		
Natural history of disease		
Disease rate		

○ **31. For which of the two types of studies are the following flaws likely to be more of a problem?** (Rate as "potential problem," "less of a problem," or if appropriate, "advantage.")

	Case-Control	Cohort
Recall bias		
Loss to follow up		
Selection bias		

○ **32. Which type of study (case-control or cohort) would you have conducted first and why? Why conduct a second study? Which type of study should this be?**

○ **33. Which of the following criteria for causality are met by the data from these two studies?**

	YES	NO
Strength of association		
Consistency with other studies		
Exposure precedes disease		
Dose-response effect		
Biologic Plausibility		