

APPENDIX 4: CASE STUDY—CIGARETTE SMOKING AND LUNG CANCER
ANSWER KEY AND INSTRUCTOR'S GUIDE

Objectives

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

1. Discuss the elements of case-control and cohort study design and identify advantages and disadvantages to both.
2. Discuss some of the biases that affect epidemiologic studies.
3. Calculate and interpret the odds ratios, rate ratios, rate differences, and attributable fractions.
4. Appreciate how the above measures of association do or do not reflect the strength of association and public health importance.
5. Discuss causal criteria presented by Hill (1965).

PART I

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

The first study was a case-control study because a source population was sampled for people who had the disease (lung cancer cases) and people who did not have the disease (controls). Cases and controls were examined for differences in the proportion of people who had been exposed to the exposure being investigated, which in this instance was cigarette smoking.

Question 1b What makes the second study a cohort study?

The second study is a cohort study because people in the source population were classified according to whether they were exposed or nonexposed. This is a prospective cohort study, since people in the cohort are followed forward in time to determine disease (lung cancer) incidence.

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

Question 2 Why were hospitals chosen as the setting for this study? What other sources of cases and controls might have been used?

The hospital is a good place to find cases. In addition, diagnoses are likely to be accurate and, once cases are identified, you have a captive audience for questioning.

Other sources of cases include cancer registries, death certificate data bases, insurance and other administrative data sources, and employment records.

Sources of controls include friends and neighbors of cases and random samples of the source population (e.g., a random-digit dialing phone survey).

Question 3 What are the advantages of selecting controls from the same hospital as cases?

Some advantages include convenience. In addition, matching on time, socioeconomic factors, place of residence, and diagnostic practices would be helpful in mitigating confounding by these factors.

Question 4a The case series are all patients admitted to some 20-odd hospitals with a new diagnosis of lung cancer. How would you define the study base (“source population”) for these cases?

The study base consists of all people in the hospitals’ catchment areas who would be hospitalized if they developed lung cancer.

Question 4b The controls were patients with other disorders treated at these same hospitals. Do you think the controls would fairly represent the study base?

Hospitalized patients are not very representative of the source population; they are more likely to have multiple chronic conditions (Berkson’s bias), and are more likely to be smokers.

Question 4c How may these issues of representativeness affect the study’s results?

The control group in a case-control study are used to estimate the relative size of the exposed and nonexposed components of the source population. Hospitalized controls may be in the hospital for a smoking-related diagnosis. Thus, the prevalence of smoking in a hospital is greater than that of the source population. The effect of a higher prevalence of smoking in the hospitalized control series is that the rate ratio in the population will be underestimated. This is a form of *selection bias*.

Question 5 [Using the data in Table A4.1] calculate the odds ratio of lung cancer associated with smoking. Include a 95% confidence interval for the odds ratio. Interpret your results.

Odds ratio estimate = $(1350)(61) / (1296)(7) = 9.08$
The 95% confidence interval for the OR = (4.14, 19.92)

Interpretation: Within the source population, the rate of lung cancer smokers is about 9 times that of nonsmokers.

Question 6 [Using the data in Table A4.2,] calculate the odds ratio associated with each level of smoking compared to the baseline provided by nonsmokers. Interpret your results. (Odds ratio 1 is for the comparison of 1–14 cigs/day to 0 cigs/day; odds ratio 2 is for the comparison of 15–24 cigs/day to 0 cigs/day; odds ratio 3 is for the comparison of 25+ cigs/day to 0 cigs/day.)

$$\hat{OR}_1 = \frac{(565)(61)}{(706)(7)} = 6.97$$

$$\hat{OR}_2 = \frac{(445)(61)}{(408)(7)} = 9.51$$

$$\hat{OR}_3 = \frac{(340)(61)}{(182)(7)} = 16.28$$

The odds ratios increase progressively, suggesting a biological gradient (dose-response relation) between the amount smoked and the strength of the association.

Question 7 While this study demonstrates a clear association between smoking and lung cancer, cause and effect are not the only possible explanation. What are other possible explanations for the association?

The association may be explained by various sources of systematic error in the study (bias) and random error (imprecision). The three types of bias are selection bias, information bias, and confounding. Imprecision is an unlikely explanation for the association seen in this large study.

Selection bias may have occurred in this study because controls were selected from the hospital. Since hospitalized people are more likely to be smokers than the source population, this would cause a bias in which the actual risk ratio associated with smoking would be underestimated.

Information bias would occur in this study if the lung cancer cases were more likely to recall their smoking history than controls. When this study was undertaken, in the 1940s, the smoking–lung cancer hypothesis was not widely known. In addition, since the controls were hospitalized as were the cases, the controls were as likely as the cases to be introspective about prior exposures. Thus, information bias is an unlikely explanation for the association.

A confounder is an extraneous factor that biases the exposure–disease relation being studied. To be confounder, the extraneous factor is normally associated with (but not a consequence) of the exposure and is an independent risk factor for the disease. Thus, the extraneous factor of older age would be a confounder if it were associated with smoking since age is definitely associated with lung cancer risk. There are many extraneous factors that could (conceivably) explain the exposure–disease relation.

PART II

Question 8 How might the response rate of 68% affect the study's results?

As a general rule, we like to see as high a response rate as possible. Realistically, a 68% response rate is considered acceptable for a mail-based study. If failure to respond is not related to both the exposure and disease, then non-response will decrease the power of the study to discern a difference but will not bias the measure of association. If non-response is related to both the exposure and disease, then the study could be biased. Thus, if possible, the researcher should characterize non-response rates in the cohorts. When the non-response rate differs in the cohorts, the researcher is alerted to a greater potential for *selection bias*.

Question 9a Table A4.3 shows the number of lung cancer deaths and person-years in smokers and nonsmokers. Calculate the lung cancer rate in each group. Then calculate the rate ratio and rate difference associated with smoking. Interpret your results.

$$R_1 \text{ (per 1000 person-years)} = 133 / 102,000 \times 1000 = 1.30$$

$$R_0 \text{ (per 1000 person-years)} = 3 / 42,800 \times 1000 = 0.07$$

$$RR = 1.30 / 0.07 = 18.6$$

$$RD \text{ (per 1000 person-years)} = 1.30 - 0.07 = 1.23$$

The rate ratio of 18.6 shows that the rate of lung cancer in smokers was 18.6 times that of nonsmokers. The rate difference of 1.23 per 1000 person-years demonstrates an excess of 1.23 cases per 1000 person-years. The rate ratio is a relative measure of effect and determines the strength of the association between the exposure and disease. The rate difference is an absolute measure of effect.

Question 9b Calculate the lung cancer rate in the entire cohort (smokers and nonsmokers combined). If no one had smoked in the cohort, we may assume that the cohort would have had the lung cancer rate of nonsmokers. What proportion of cases in the cohort would have been averted if no one had smoked? What is this fraction called?

$$\text{Recall } R_0 \text{ (per 1000 person-years)} = 0.07$$

$$\text{Calculate rate in the entire cohort: } R \text{ (per 1000 person-years)} = 136 / 144,800 \times 1000 = 0.94$$

The question asks for the population attributable fraction: $AF_p = 0.94 - 0.07 / 0.94 = 0.925$. This suggests 92.5% of the cases in the cohort could have been averted if no one smoked.

Question 10 Table A4.4 lists the number of lung cancer deaths by amount smoked. Compute the rate ratio for each smoking category, using the nonsmokers as the baseline rate. Interpret your results.

Cigarettes per day	Lung Cancer Cases	Person-Years	Rate per 1000 person-years	Rate Ratio	Rate Dif. per 1000 person-years
25+	57	25,100	2.27	32.4	2.20
15 - 24	54	38,900	1.39	19.8	1.32
1 - 14	22	38,600	0.57	8.1	0.50
0	3	42,800	0.07	referent	referent

Notice the dose-response relation between amount smoked and lung cancer rates.

Question 11 Table A4.5 lists lung cancer mortality rates according to the duration of smoking cessation in people who had quit smoking. What do these data say about smoking cessation?

The lowest rate is seen in those who never smoked. However, rates decrease with the time since persons last smoked, although even after 20 years of not smoking the rate is nearly three times greater than in never-smokers.

Question 12a The cohort study also provided information about cardiovascular mortality rates. Table A4.6 presents some of this data alongside lung cancer mortality data. Which disease, cardiovascular disease or lung cancer, has a stronger association with smoking?

Rate ratios indicate a much stronger association between smoking and lung cancer mortality than between

smoking and cardiovascular disease mortality (18.5 vs. 1.3).

Question 12b If the rate difference is used as an index of the effect of smoking per 1000 person-years of exposure, on which disease, cardiovascular disease or lung cancer, does smoking have a greater effect?

Rate difference indicate greater effect of smoking for cardiovascular disease (2.19 per 1000 person-years vs. 1.23 per 1000 person-years)

Question 13 Odds ratios from the case-control study and rate ratios from the cohort study are listed side by side in Table A4.7. How do these results compare? Can you suggest a plausible explanation for these discrepancies?

The odds ratios in the case-control studies are consistently smaller than the rate ratios. An explanation for this discrepancy is related to the use of hospitalized controls in the case-control study. As discussed earlier, use of hospitalized controls was likely to result in a bias that tended to underestimate the risk associated with smoking.

Question 14a Using Table A4.8, check the box corresponding to the advantage held by the case-control or cohort study. For example, since case-control studies require smaller sample sizes than cohort studies, this box is checked under the case-control column.

	Case-Control	Cohort
Allows for smaller sample size	X	
Costs less	X	
Shorter study time	X	
Better suited to study rare disease	X	
Better suited to study rare exposure		X
Convenient when studying multiple exposures	X	
Convenient when studying multiple diseases		X
Able to study natural history of disease		X
Can estimate disease rates		X
Less prone to selection biases		X
Less prone to recall biases		X
Less prone to loss to follow-up biases	X	

Question 14b Why was the case-control study done before the cohort study?

The case-control study was quicker, easier, and less expensive. The cohort study was more difficult and

expensive to complete, and was slower to yield results. However, completion of the cohort study provided confirmation of the findings from the case–control study and allowed for direct calculation of rates and more intuitive interpretation of results.

Question 15 In your opinion, which of the criteria for causality have been met by evidence presented in *this* case study?

Strength of association:	Yes
Consistency between studies:	Yes
Temporal sequence:	Yes
Biological gradient:	Yes
Specificity of effect:	No
Biological plausibility:	No