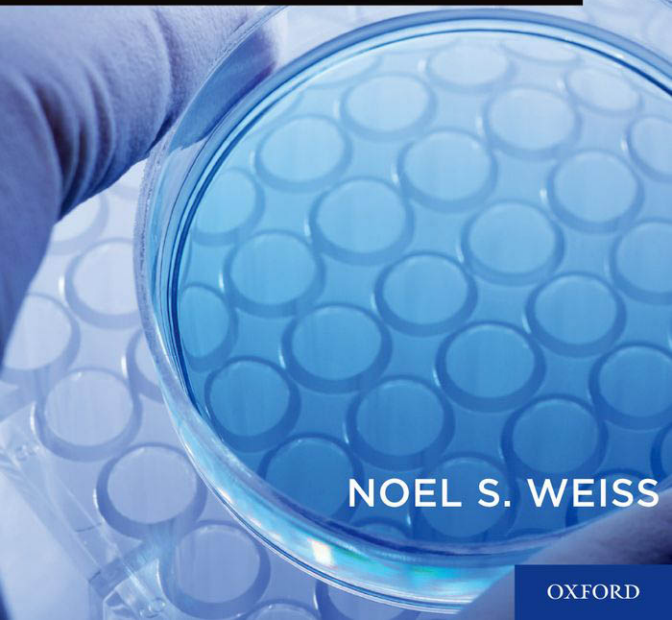
EXERCISES IN EPIDEMIOLOGY

APPLYING PRINCIPLES AND METHODS



NOEL S. WEISS

OXFORD

EXERCISES IN EPIDEMIOLOGY

This page intentionally left blank

EXERCISES IN EPIDEMIOLOGY

Applying Principles and Methods

NOEL S. WEISS

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Oxford University Press, Inc., publishes works that further
Oxford University's objective of excellence
in research, scholarship, and education.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Published by Oxford University Press, Inc.
198 Madison Avenue, New York, New York 10016
www.oup.com

Oxford is a registered trademark of Oxford University Press

Library of Congress Cataloging-in-Publication Data

Weiss, Noel S., 1941-
Exercises in epidemiology : applying principles and
methods / Noel S. Weiss.
p. ; cm.
Includes bibliographical references and index.
ISBN 978-0-19-979678-6
1. Epidemiology—Problems, exercises, etc. I. Title.
[DNLM: 1. Epidemiologic Methods—Problems and Exercises. WA 18.2]
RA652.7.W45 2012
614.4—dc23
2011017403

---

9 8 7 6 5 4 3 2 1
Printed in the United States of America
on acid-free paper

CONTENTS

This page intentionally left blank

INTRODUCTION

There are a *lot* of texts that deal with the principles and methods of epidemiology. I've been a coauthor of one of these myself. All of the texts, to a greater or lesser extent, provide examples of real or hypothetical epidemiologic studies to illustrate a given principle or method. For many (probably most) readers of these books, the examples help to solidify an understanding of the topic at hand.

What the examples do not provide is the opportunity to consider, on one's own, how a particular issue ought to be dealt with, or how a particular question should be addressed. The purpose of this book is to supplement the material contained in the textbooks in such a way that the reader is forced to: (1) identify situations in which the validity or accuracy of a particular design or analytic approach may be limited; and (2) determine how that limitation might be overcome. Such actions are just those that epidemiologists have to take when they are planning research or are reviewing that of others.

The key word in the preceding paragraph is *supplement*. The present book cannot stand alone as a means of learning about epidemiology, or even as a means of being introduced to the

subject. My hope is that the exercises contained in it can extend the knowledge of students of epidemiology, and equip them more fully to deal with the real world problems and issues that they'll encounter in their professional lives.

The book is organized into seven chapters, each of which contains a set of questions and answers to those questions. Any reader who believes a given answer is incomplete (or wrong!) is welcome to communicate with me (nweiss@uw.edu). With the reader's permission, his/her suggestions will be posted on the "discussion" page of that section of the Oxford University Press website devoted to this book—perhaps with an additional comment from me. No doubt all readers of this portion of the website will benefit from learning of different viewpoints.

To minimize the likelihood of an ambiguous question being present in this book, or an incomplete or incorrect answer, I enlisted the help of the following persons to review parts of the draft manuscript: Peter Cummings, Paul Doria-Rose, Sarah Lowry, Amanda Phipps, Gaia Pocobelli, and Ali Rowhani-Rahbar. The contributions of each of them helped to make the chapters of the book that you are reading better than the draft chapters that they received from me.

EXERCISES IN EPIDEMIOLOGY

This page intentionally left blank

CHAPTER 1

Rates and Proportions

EPIDEMIOLOGISTS LOVE denominators. Sometimes we divide the number of numerator events among exposed individuals by the total number of exposed individuals, so that we can calculate the *proportion* of (say) eaters of potato salad at a picnic who were diagnosed with a Salmonella infection during the ensuing 48 hours. At other times, we use a person-time denominator enabling us to calculate the *rate* of (say) lung cancer in persons who have been employed in a given industry. Depending on the question being addressed, we may seek to estimate a proportion or a rate. The accuracy of that proportion or that rate will depend on our ability to measure correctly both numerator and denominator.

This page intentionally left blank

Question 1.1 A recent study observed that 1 in 20 persons with cancer later were diagnosed with a second cancer. In the general population, the lifetime probability of being diagnosed with cancer is considerably greater. Is this evidence of immunity developed as a result of the first cancer?

Answer 1.1 This is not necessarily evidence of immunity. What's not being taken into account is the very different denominator for each of the two groups—the amount of person-time at risk. Among cancer patients, person-time begins to accrue as of the date of diagnosis, typically in mid- to late life. In the general population, person-time begins to accrue at birth.

Question 1.2 A study on ovarian cancer observed the following pattern of histologic type and race among its participants.

	Type of tumor		
Race	*Mucinous*	*Other*	*Total*
Caucasian	33 (13%)	225 (87%)	258 (100%)
Asian	55 (27%)	151 (73%)	206 (100%)

The authors concluded that Asian women "had a higher incidence of mucinous tumors" than did Caucasian women. What reservations do you have regarding this conclusion?

Answer 1.2 The observed proportional distribution of histologic type by race could be due to a relatively high incidence rate of mucinous tumors in Asian women, or as well to Asian women having a low rate of other ovarian tumors. For example, the rates below would give rise to the numbers presented in this question:

	Type of tumor (rate per 100,000 woman-years)		
Race	*Mucinous*	*Other*	*Total*
Caucasian	1	6.7	7.7
Asian	1	2.7	3.7

Question 1.3 The following is excerpted from a news item in the *British Medical Journal*:

> The clinical features of more than 1000 patients with lung cancer presenting to 46 UK hospitals have been analyzed. The results showed that women under 65 are particularly at risk of small cell lung cancer—34% presented with this form of the disease compared to 18% of men.

Assume that: (1) the distribution of histologic types of lung cancer in the patients under 65 years in the 46 UK hospitals accurately reflects that of all U.K. lung cancer patients; and (2) the difference between the figure of 34% in women and 18% in men is not due to chance. Under what circumstance could the observed difference *not* be indicative of a difference in the incidence of small cell lung cancer between U.K. men and women under 65 years?

Answer 1.3 The proportional incidence by gender will not be an indication of the absolute incidence if the incidence of non–small cell lung cancer is different in men and women.

For example:

	Men	*Women*
Small cell cancers (%)	360 (18%)	340 (34%)
Other types	*1640*	*660*
Total	2000	1000

In the above example, assuming the numbers of men and women in the population are similar to one another, the rate of small cell lung cancer by gender is nearly identical. The disparity in the *proportional* incidence comes from the disparity in the rates of lung cancer that are not of the small cell type.

Question 1.4 The following statement appeared in a review article:

> In 1996 in the United States, a total of about 34,000 new cases of endometrial cancer occurred, as well as approximately 6,000 deaths from this disease. The case-fatality is approximately 28%.

a. Assuming that the data described in the first sentence are correct, why is it unlikely that the case-fatality from endometrial cancer is truly 28%?
b. Describe a circumstance under which the data in the first sentence *and* a case-fatality of 28% for endometrial cancer in U.S. women could both be true.

Answer 1.4

a. If "equilibrium" exists—in other words, no change in number of cases or case fatality over time—the case fatality among women with endometrial cancer should be 6,000/34,000 = 18%.
b. The 6,000 deaths in 1996 occurred primarily in women diagnosed with endometrial cancer prior to that year. So, if the incidence of endometrial cancer had very recently increased to a large degree, the appropriate denominator for the calculation of case-fatality would be a number much smaller than 34,000 (specifically, 21,429 cases in order to generate a case-fatality of 28%).

Question 1.5 Let's say you've conducted a cohort study to determine some long-term consequences of surgical treatment of patients with cataracts. For 174 patients who underwent surgery and 103 other patients with cataracts who did not, you've used records of the state department of motor vehicles to determine who has been involved in a motor vehicle crash as a driver. From periodic interviews with study subjects, you are able to estimate the number of miles each one has driven during a 2-year follow-up period.

The results of the study are as follows:

Study group	*No. of persons*	*No. of crashes*	*No. of miles driven*
Surgery	174	27	5,677,867
No surgery	103	23	2,569,639

Assume the two groups of patients are exactly comparable with respect to baseline characteristics that predict automobile crash occurrence, including driving behavior, and that no misclassification is present in the study.

a. Estimate the influence of cataract surgery on crash rate while driving.
b. Estimate the overall influence of cataract surgery on the risk of an automobile crash, in other words, that which would include a possible influence of the surgery on driving behavior.

Provide the rationale for your answers.

Answer 1.5

a. The answer is the relative rate based on the number of miles driven

$$\frac{27/5{,}677{,}867}{23/2{,}569{,}639} = 0.53$$ – since it allows for the number of driver-miles at risk.

b. Since the number of miles driven seems to have been influenced by the receipt of surgery, the assessment of the aggregate impact should *not* consider this, and the relative risk of 0.69 $\left(\frac{27/174}{23/103}\right)$ should be used.

Question 1.6 The following is paraphrased from an article in the *British Medical Journal*[1]:

> Although the relative rate of myocardial infarction associated with cigarette smoking is higher in women than in men, smoking may well cause a higher rate of myocardial infraction in men who smoke than in women who smoke.

Under what circumstance could this be true?

Answer 1.6 It could be true if, among nonsmokers, the incidence of myocardial infarction (MI) in men were higher than that in women. For example, assume that in a certain age group the annual incidence of MI was 3 per 1,000 in men and 1 per 1,000 in women. Among men, a relative rate of 2 associated with smoking would produce a rate difference of (2*3/1,000)–3/1,000 = 3/1,000 person-years. Among women, a higher relative rate–3–would produce a rate difference that is smaller than this: (3*1/1,000)–1/1,000 = 2/1,000 person-years.

Question 1.7 The following is excerpted from a letter to the editor of a medical journal:

> We have observed renal cell carcinomas in 6 out of 412 patients with analgesic nephropathy (1.4%), treated over the past 12 years. The incidence of renal cell carcinoma in the general population is 7.5 per 100,000 population per year, so the prevalence found in patients with analgesic nephropathy is highly significant ($p < .005$).

What additional information on these patients with analgesic nephropathy would be needed in order to better assess the possibility that they are at increased risk of renal cell carcinoma?

Answer 1.7 At the very least, we would need the age-specific person-time at risk for the diagnosis of cancer among participants with analgesic nephropathy. This would permit a comparison of age-adjusted *rates* of renal cell carcinoma between these patients and the general population.

Question 1.8 The Second National Health and Nutrition Examination Survey was a cross-sectional survey conducted from February 1976 to February 1980, with a probability sample of 27,801 persons in the United States.

The following table presents some data obtained in the survey:

Percentage of children 6 months through 4 years with a history of eating unusual substances by selected characteristics: United States, 1976–1980

	No. examined	*Percentage with history of eating unusual substances*
Blood lead level in micrograms per deciliter:		
30 or more	117	16.2
20–29	503	14.1
Less than 20	1,752	5.2

Earlier studies have shown that elevated blood lead levels (30 μg/dl or higher) are associated with slowed intellectual development in children. At issue in the present analysis is whether eating "unusual" substances (e.g., paint) contributes to elevated blood lead levels.

In U.S. children 6 months through 4 years of age, can you determine the likelihood of having a blood lead level of >30 μg/dl for those with a history of eating unusual substances relative to the likelihood for those with no such history? If yes, what is it? If no, why not?

Answer 1.8

History of eating unusual substances	*Blood lead level (microg/dl)*		
	≥30	*<30*	*All*
Yes	19	162	181
No	98	2,093	2,191
			2,372

$$\text{Relative risk} = \frac{19/181}{98/2{,}191} = 2.3$$

Question 1.9 The following data were obtained in a very large cohort study conducted in Korea during 1993–2002 that examined potential risk factors (including the prevalence of hepatitis B surface antigen positivity (HbsAg+)) for mortality from hepatocellular carcinoma (HCC).

	No. of HCC deaths	*Rate per 100,000 person-years*
Men		
HbsAg+	1522	405.2
HbsAg–	734	21.8
Women		
HbsAg+	37	58.4
HbsAg–	9	1.2

a. For men and women, separately, estimate the relative mortality from HCC associated with being HbsAg+, and also the mortality difference.
b. One of the above measures of excess mortality is greater in men; the other greater in women. How can this be?

Answer 1.9

a. Relative mortality

Men: $405.2/21.8 = 18.6$
Women: $58.4/1.2 = 48.7$

Mortality difference (per 100,000 person-years)

Men: $405.2 - 21.8 = 383.4$
Women: $58.4 - 1.2 = 57.2$.

b. The annual mortality from HCC, in the absence of active infection with hepatitis B, differs greatly by sex: 21.8 per 10^5 for men versus 1.2 per 10^5 for women. Thus, an absolute increase in mortality of 57.2 per 10^5 experienced by Korean women is very large in relative terms (relative mortality = 48.7). In men, the larger absolute mortality difference (383.4 per 10^5) is not nearly so large on a ratio scale, since it is superimposed not on a "baseline" mortality rate of 1.2 per 10^5, but on the higher male "baseline" rate of 21.8 per 10^5.

Question 1.10 Black men in the United States have a substantially higher incidence of prostate cancer than U.S. white men. Let's say there's a variant of the androgen receptor gene that's more common in black than white men in the United States—50% versus 30%—that is also associated with a doubling of incidence of prostate cancer in American men of either race.

What would be the relative incidence of prostate cancer, black versus white American men, if the genetic marker were the sole risk factor for this disease that differed between the two races?

Answer 1.10 If x = incidence of prostate cancer in men without the variant genotype, the incidence of prostate cancer in white men would be a weighted average of the incidence in the 70% of men without the variant genotype and the 30% who have it: $7x + .3(2x) = 1.3x$. The incidence in black men would be $.5x + .5(2x) = 1.5x$, because half have the variant genotype and half do not. If, in terms of prostate cancer risk, white and black men were identical save for the prevalence of this genotype, black men would have an incidence that was $1.5x/1.3x = 1.15$ times that of white men.

Question 1.11 You read a magazine article in which a medical columnist has expressed concern that the mean age at which colorectal cancer is diagnosed among Americans who smoke cigarettes and consume alcohol is lower than among their fellow citizens who neither smoke nor drink. Assume that the age distribution is the same between Americans who smoke and drink and those who do not. Must it be true that, among relatively young American adults, the incidence of colorectal cancer is higher in cigarette smokers/alcohol drinkers than in other persons? If yes, why? If not, why not?

Answer 1.11 No. For example, if among *older* persons the incidence of colorectal cancer were relatively low in those who smoked and consumed alcohol, with the incidence among younger persons who smoked and consumed alcohol being the same as that of young abstainers, the mean age of diagnosis of smokers/drinkers also would be lower than that of abstainers.

Question 1.12 The following statement was made in a newspaper article that sought to provide data bearing on the efficacy of seat belts in preventing deaths that occur in automobile crashes:

> Of the 649 people who died in traffic accidents in Washington last year, 55 percent were not wearing seat belts. In those same fatal crashes, 73 percent of people who were belted in survived without serious injury.

Does this statement support the hypothesis that seat belt use saves lives? Explain.

Answer 1.12

	Dead	*Alive*
Belt	45%	?
No belt	55%	?
Total	100%	100%

The data do not bear on the hypothesis. What is needed instead is information on the percentage of persons who survived these crashes who were unbelted. There would be evidence of efficacy to the extent that this figure was smaller than 55%. Alternatively, one could compare the percentage of *un*belted persons who survived without serious injury to the figure of 73% for belted individuals:

	Dead	*Alive*	*Total*
Belt	27%	73%	100%
No belt	?	?	100%

Question 1.13 During a recent decade in the United States, the annual proportion of all women ages 25 to 29 years who gave birth to a first child rose from .031 to .039. In this same decade, however, the annual proportion of childless women ages 25 to 29 years who gave birth to a child fell from .114 to .092. How is it possible that these two trends can be in opposite directions?

Answer 1.13 The numerator for the two proportions is the same, in other words, the annual number of 25- to 29-year-old American women who gave birth to their first child. But the denominator for the second proportion—the number of 25- to 29-year-old childless women—is but a part of the first denominator (all 25- to 29-year-old women). In order for the incidence of first births to have risen overall but to have declined among childless women, it must be true that the fraction of 25- to 29-year-old women who were childless must have risen during the decade. This more than compensated for the declining first-birth incidence in childless women and caused a rise in the first-birth incidence in 25- to 29-year-olds as a whole.

Question 1.14 The rate of suicide among American physicians, relative to the corresponding rate in the population as a whole, varies by gender. Among men, the rate in physicians is 1.5 times higher, whereas among women the corresponding relative rate is 3.0. It turns out that the rate of suicide in American male and female physicians is identical. For American men and women in general, what is the relative rate of suicide in men compared to women?

Answer 1.14

Pm = rate of suicide in male physicians,
Pf = rate of suicide in female physicians.
M = rate of suicide in American men
W = rate of suicide in American women

$$RR, \text{men} = \frac{Pm}{M} = 1.5$$

$$M = \frac{Pm}{1.5}$$

So, for American men in general, their rate of suicide is that of the male physicians divided by 1.5.

$$RR, \text{women} = \frac{Pf}{W} = 3.0$$

$$W = \frac{Pf}{3.0}$$

Similarly, American women have but one-third the rate of suicide of female physicians.

Now, because Pm and Pf are the same (we'll label this rate as P),

$$\frac{M}{W} = \frac{P/1.5}{P/3.0} = \frac{3.0}{1.5} = 2.$$

American men, as a whole, have twice the rate of suicide as American women.

Question 1.15 A study of suicide among men with cancer was conducted in the United States.[2] The goal of the study was to enable health professionals to "be aware of the potential for suicide in cancer patients." Some of the site-specific data are presented below.

Type of cancer	*No. of men with cancer*	*No. of suicides*	*Suicides per 100 men (95% CI)*
Lung	102,940	215	0.21 (0.18–0.24)
Melanoma	19,377	46	0.24 (0.17–0.32)
Thyroid	5,339	14	0.26 (0.14–0.42)

It had been hypothesized that the risk of suicide during any given period of time following diagnosis would be greatest for types of cancer with a poor prognosis (e.g., lung) than types with a good prognosis (e.g., melanoma, thyroid). Do the above data argue against this hypothesis? (Assume that the distribution of demographic characteristics bearing on suicide occurrence is similar across the three types of cancer.)

Answer 1.15 The data do *not* argue against the hypothesis. The analysis fails to consider person-time at risk. Because this is, on average, considerably greater for a man with melanoma or thyroid cancer than for a man with lung cancer, the *rate* of suicide (i.e., number of suicides divided by person-time at risk) in the latter group must be higher than the rate for the other two groups.

Question 1.16 In a study of oral cancer, you observe that 17% of the Hispanic cases are younger than 40 years, as compared to 4.8% of non-Hispanic men with oral cancer ($p < .05$).

Assume that the ascertainment of cases of oral cancer was equally complete in the Hispanic and non-Hispanic men, and that the above difference was not due to chance. Does this finding necessarily imply that in the population under study the risk of developing oral cancer is elevated in Hispanic men under 40 years of age compared to non-Hispanic men of similar age? Explain.

Answer 1.16 No. The age disparity could simply be a reflection of the relatively younger age of Hispanic males in the population under study.

For example:

	Hispanic men			*Other men*		
Age (years)	*No. of cases*	*Person-years*	*Incidence per 100,000*	*No. of cases*	*Person-years*	*Incidence per 100,000*
<40	17	100,000	17	17	100,000	17
≥40	83	100,000	83	337	406,000	83
% under 40	17%			4.8%		

Or, beyond this, a high proportional incidence of oral cancer in younger Hispanic men could be due to an atypically low absolute incidence in *older* Hispanic men.

For example:

	Hispanic men			*Other men*		
Age (years)	*No. of cases*	*Person-years*	*Incidence per 100,000*	*No. of cases*	*Person-years*	*Incidence per 100,000*
<40	17	100,000	17	17	100,000	17
≥40	83	100,000	83	337	100,000	337
% under 40	17%			4.8%		

Question 1.17 A population-based case-control study of Guillain-Barré syndrome (a neurological disease) conducted in 1992–1994 in 4 states estimated the risk of this disease to be 1.7 times greater among adults who had received influenza vaccine in the prior 6 weeks than those who had not. The investigators also estimated that the added risk of Guillain-Barré syndrome associated with the receipt of influenza vaccine was about one per million persons during the first 6 weeks after vaccination.

From these data, can you calculate the 6-week incidence among adults in the 4-state population who did *not* receive the vaccine? If yes, what is that incidence? If no, why not?

Answer 1.17 The difference of 1 per million in the 6-week incidence between persons who did (I_e) and did not (I_o) receive the vaccine is $I_e - I_o = 1.7I_o - I_o$. Therefore, $I_o = 1/0.7 = 1.43$ per million.

Question 1.18 The incidence of stomach cancer in country *X* is 8.0 per 100,000 per year. The incidence rate in nearby country *Y*, with a similar age-sex-race composition as country *X*, is 10.0. You are concerned with explaining this difference. You know that 5% of people in country *Y* drink tea containing suspected carcinogen *A*, whereas nobody in country *X* drinks this tea. In order for this to be the sole explanation of the difference in the incidence rates of stomach cancer between the two countries, how strongly must carcinogen-*A*-tea drinking be associated with stomach cancer?

Answer 1.18 If all the difference were due to ingestion of carcinogen *A* in tea, the incidence of stomach cancer in country *Y* could be described as follows:

$$10 = .95(8) + .05t$$

where t = incidence in drinkers of *A*-containing tea

$$t = \frac{10 - .95(8)}{.05} = 48$$

$$48/8 = 6$$

Therefore, a relative risk of 6 (associated with drinking *A*-containing tea) is required for the whole of the difference in rates between the two countries to be attributable to this exposure.

Question 1.19 During 1993–2001, men at 10 U.S. study centers were invited at random to receive annual PSA screening for 6 years (plus annual digital rectal exams for 4 years; n = 38,343) or no intervention (n = 38,350).[3] Through 10 years from the time of randomization, there were 3,452 cases of prostate cancer diagnosed and 83 deaths from this disease in the group invited for screening, versus 2,974 cases and 75 deaths from prostate cancer among men in the control arm. The investigators noted that, among men diagnosed with prostate cancer in the screening and control arms of the trial, there were 312 and 225 deaths from other causes, respectively, a difference of 87 deaths. They went on to speculate that his latter difference "was possibly associated with over-diagnosis of prostate cancer."

What would be a better approach to quantifying the likelihood of death from causes other than prostate cancer between men who were invited and those who were not invited to be screened?

Answer 1.19 Among men diagnosed with prostate cancer, ideally the *rate* of death from causes other than prostate cancer, not simply the number of such deaths, would be compared: that is, the number of deaths divided by the number of person-years. Even though the sizes of the groups invited and not invited to be screened were nearly identical, the number of men diagnosed with prostate cancer in the former group was larger (by 16%, 3452 vs. 2974), because screening identified a number of malignancies that otherwise would not have been diagnosed during the follow-up period. Failure to take into account the relatively larger number of person-years in the invited men diagnosed with prostate cancer would lead to a falsely high mortality rate in that group.

To the extent that the age distribution of men with screen-detected prostate cancer differed from that of men whose cancer was diagnosed for other reasons, a valid comparison would require age adjustment as well (given the strong association between age and mortality rates).

CHAPTER **2**

Causal Inference

EPIDEMIOLOGISTS' PRIMARY responsibility to society is to provide data relevant to the prevention of disease and injury. We do this by examining the association between disease or injury and potential etiologic factors. When across all relevant epidemiologic studies there does appear to be an association, it is necessary to infer whether that represents cause and effect—preventive measures would be appropriate only when a judgment of cause and effect can be made. To a great extent, the process of distinguishing causal and noncausal associations is subjective: persons may disagree on the relative importance of the various elements that go into such an inference, and on the degree to which those elements are present. However, subjective or not, we cannot escape the need to try to draw causal inferences—an efficient program of prevention depends on our success in correctly distinguishing associations that are causal from those that are not.

This page intentionally left blank

Question 2.1 The following is excerpted from an abstract of an article on sudden infant death syndrome[4]:

> *Objective:* To assess the role of parental bedsharing in sudden infant death syndrome (SIDS)-like deaths, this study examines the hypothesis that, compared with other SIDS cases, the age distribution of deaths associated with bedsharing should be lower.
>
> *Methods:* For 84 SIDS cases in Cleveland, Ohio, 1992 to 1996, age at death, maternal weight, and other risk factors for SIDS were compared for cases grouped according to bedsharing status.
>
> *Results:* Mean ages at death were 9.1 weeks for 30 bedsharing and 12.7 for 54 nonbedsharing cases.
>
> *Conclusion:* By demonstrating that among an urban population at high risk for SIDS, bedsharing is strongly associated with a younger age at death, independent of any other factors, this study provides evidence of a relationship between some SIDS-like deaths and parent-infant bedsharing.

In this study, assume that information on bedsharing was completely accurate. What is your primary reservation regarding the authors' assertion that their study "provides evidence of a relationship between some SIDS-like deaths and parental bedsharing"?

Answer 2.1 Parental bedsharing with an infant may simply be an age-related phenomenon: the younger the infant, the more likely the parents are to share a bed with him/her. If so, then bedsharing may have no etiologic relevance to SIDS at any age.

Question 2.2 Quoted from a magazine of an insurance agency:

> Does your automobile have body damage? Fix it, and you'll significantly reduce your probability of involvement in another traffic mishap. There is a distinct psychological advantage to having even minor auto collision damage repaired as soon as possible. Studies have shown that drivers of newly repaired automobiles tend to drive more defensively than those with unrepaired damage.

Assume that the studies referred to are cohort (follow-up) in type and that they have shown unequivocal differences in the rate of second accidents between persons who did and did not repair the damage resulting from an initial accident. What is your main reservation concerning the conclusion that repairing auto damage influences driving behavior?

Answer 2.2 The association may be due to a third (confounding) factor—persons who choose to have damaged automobiles repaired may drive more defensively than persons who choose not to do so, no matter what the condition of their cars.

Question 2.3 For some time, women taking SSRIs (a class of antidepressant drugs) during pregnancy have been encouraged to stop this medication 14 days prior to the end of the pregnancy, due to concerns about a possible deleterious impact on labor and delivery. A study conducted to examine this issue observed an increased risk of prematurity (<37 weeks gestation) in infants born to mothers who had taken SSRIs and had *not* stopped taking them 14 or more days prior to delivery (women not taking SSRIs during their pregnancy served as the basis for comparison). No corresponding association was seen for discontinuation of SSRIs at least 14 days before delivery.

Even though you have no reason to be concerned with misclassification of exposure or outcome status in this study, or with a difference in risk factors for prematurity at the start of pregnancy between the three categories of women (nonusers of SSRIs, users within the last two weeks of pregnancy, users only prior to the last two weeks of pregnancy), you are concerned that the study does not provide a valid result. What is the basis for your concern?

Answer 2.3 These data may be indicative of a causal relation, but one in the opposite direction. That is, premature delivery in a woman using an SSRI during pregnancy is likely to prevent her from stopping the drug, rather than the reverse.

Question 2.4 The following question pertains to a news item that appeared in a medical journal:

> People screened for lung cancer by spiral CT have accelerated and prolonged quit rates of smoking, regardless of whether the screening shows disease. Researchers found that 1 year after scanning, 14% of smokers had stopped smoking; by contrast, the rate among the general smoking population was 5–7%. The findings suggest that screening is an ideal place to provide cessation messages, say the researchers.

Assume that the difference between the figures of 14% and 5–7% is not due to chance and is not due to differences in demographic characteristics between persons who do and do not receive spiral CT screening. What is an explanation for this difference apart from a genuine impact of attending the screening program on the likelihood of smoking cessation?

Answer 2.4 Due to the likely presence of a great deal of confounding, the data obtained in this study do not have the ability to assess the efficacy of cessation messages to persons who smoke cigarettes. It could be that those smokers who are sufficiently health conscious to receive screening for lung cancer are overrepresented with persons who, apart from any additional health education or advice, are likely to quit during the coming year.

Question 2.5 The following quotation comes from a review of research on diet in relation to levels of serum lipids:

> Many (studies) were limited by having a small number of participants. Diet studies quoted to this day as authoritative had as few as five subjects. If any enduring truth has emerged about human beings and diet, it is that everyone is remarkably different, and studies that don't involve dozens if not hundreds of participants are of extremely limited value.

How is it possible for an epidemiologic study (of any question) that involves fewer than dozens or hundreds of participants to be of more than limited value?

Answer 2.5 I had a professor of neuroanatomy who groused one day in class about a manuscript of his that was not accepted for publication, apparently because the results contained in it were based on but two dogs. He said that if he had just one dog, but could teach it to play the violin, there ought to be no concern with the sample size! If a study has identified a *strong* association (and otherwise has been designed in a way to provide a valid result), it need not be large to document this in a convincing way. For example, it took just 8 cases of vaginal adenocarcinoma and 32 controls to identify an unequivocal association with maternal use of diethylstilbestrol (7 cases and 0 controls had been exposed in utero).

Of course, when associations are not so strong—perhaps in the case of diet in relation to serum lipids—studies of dozens if not hundreds (if not more) participants are needed to document their presence and/or size.

Question 2.6 A study investigated the relationship between a woman's age at first birth and the presence of depression. The study used data from the U.S. National Maternal and Infant Health Survey, a National Center for Health Statistics-sponsored nationally representative survey that selected a stratified systematic sample of 1988 live births from state vital statistics records via a multistage cluster design. Respondents were interviewed in person an average of 17 months after delivery, and the data collected included (among other things) age, race, and the respondent's extent of depressive symptoms in the 4 weeks prior to the interview. On the scale that was used, a score of 16 or higher (out of a possible 60) was the criterion for categorizing a woman as "depressed." This analysis was limited to women who had just delivered their first child.

The following table was presented:

Age, years	*% depressed*	*Odds ratio (95% CI)**
	African-American women	
15–17	48.1	2.7 (1.9, 3.9)
18–19	36.8	1.7 (1.2, 2.5)
25–34	25.3	1.0 (reference)
	White women	
15–17	27.8	2.4 (1.4, 4.1)
18–19	32.9	3.0 (2.0, 4.7)
25–34	13.8	1.0 (reference)

*Adjusted for all relevant confounding variables.

Despite the strong association present in this study, you are reluctant to conclude that giving birth to a child as a teenager is a cause of depression. Why is this?

Answer 2.6

a. The prevalence of depression could be higher in teenagers than in 25- to 34-year-old women even in those without a recent first birth. Thus, the observed association could be related to age per se and not to having had a child at a given age.

and/or

b. Antecedent depression could be more common in women who as teenagers attempt to conceive for the first time (or who fail to prevent conception) than in women who do so at ages 25 to 34.

Question 2.7 In a study conducted among members of a prepaid health-care plan, an investigator observed that infants of women who had obtained analgesic *X* from the plan's pharmacy sometime in the year prior to delivery had twice the risk of certain major congenital anomalies compared to the infants of women who did not obtain that drug. Additionally, pregnancies in women who had obtained analgesic *X* ended in spontaneous abortion 1.8 times more commonly than did pregnancies in women who had not.

The study's findings of increased risk were subsequently criticized because of questions about the accuracy with which infants were classified as exposed or unexposed. Because analgesic *X* is relatively inexpensive and widely available without prescription, one critic argued that it was likely that some members purchased it outside the health plan. Another critic noted that there was no evidence that the analgesics obtained were actually used during pregnancy. In your judgment, could these factors account for the study's findings, in the absence of a true association? Explain.

Answer 2.7 No. The information bias resulting from these factors would be expected to be nondifferential with regard to the occurrence of a congenital anomaly or spontaneous abortion. This would result in the observed association being spuriously close to the null, rather than away from it.

Question 2.8 A study was conducted among women with either of two types of breast cancer: (1) negative for both estrogen receptor (ER) and progesterone receptor (PR); and (2) all other types combined.[5] The two groups were compared with regard to the use of a statin medication prior to diagnosis. The results were as follows:

	Type of breast cancer	
Stat in use	*ER–/PR–*	*All other*
Yes	34 (11%)	269 (15%)
No	340	1498
	374	1767

a. The difference in prior use of a statin between the two groups of cases was unlikely to be the result of chance (p = .02). The study investigators concluded that statin use could well be related to a reduced incidence of ER–/PR– breast cancer. Assuming that no confounding is present, what is an alternative interpretation of these results?

b. One analysis focused on use of lipophilic statins in relation to type of breast cancer, because of the investigators' concern that prior studies had not accounted for "the confounding effect of combining lipophilic and hydrophilic statins" when examining this question. Assuming that the large majority of women use just one type of statin (lipophilic or hydrophilic), it is likely that the investigators concern was not actually with confounding in those earlier studies, but with another potential source of bias. Which one? Explain?

Answer 2.8

a. It is possible that statin use increases the incidence of breast cancer that is not ER–/PR–.
b. The investigators' concern appears to be with exposure heterogeneity, specifically that an association between the proportional incidence of ER–/PR– breast cancer and lipophilic statin use could be obscured by having included users of hydrophilic statins in the "exposed" category.

Question 2.9 The following data were obtained in a study to evaluate venous thromboembolism risk among women after air travel.

Air travel	*Oral contraceptives*	*# of cases*	*# of controls*	*Odds ratio*	*95% CI*
No	No	54	94	1.0	Ref.
No	Yes	95	48	3.5	2.1–5.8
Yes	No	4	5	1.4	0.3–6.8*
Yes	Yes	20	2	17.4	3.9–157.0*

*Exact confidence limits.

The authors noted that they found a marked increase in risk of venous thromboembolism among women who used oral contraceptives and also had recent exposure to air travel, but argued for a cautious interpretation of this finding because of the small number of controls who reported recent air travel and oral contraceptive use. Do you agree that the small number of controls with both exposures argues for a cautious interpretation? Explain.

Answer 2.9 No. Because the lower limit of the confidence interval greatly exceeds 1, chance is an unlikely explanation for the association, notwithstanding the small number of controls with both oral contraceptive use and recent air travel.

Question 2.10 In the United States in 1998, all enriched grains and cereals were required to contain 140 micrograms of folic acid per 100g of grain. Median levels of blood folate among American women of childbearing age rose from 4.8 ng/ml in 1994 to 13.0 ng/ml in 2000.

One of the reasons for the folic acid fortification program was the results of studies (both randomized and nonrandomized) conducted prior to 1998 that documented a large increase in risk of neural tube defects in the offspring of women with low folic acid intake in the year before becoming pregnant and/or blood levels of folic acid at the start of pregnancy. However, a large, multisite, case-control study of neural tube defects in U.S. children born during 1998–2003 failed to identify a difference between mothers of cases and controls with regard to folic acid intake in the year preceding the pregnancy.[6]

Assume that the 1998–2003 study provided a valid result, and that the confidence limits around the risk estimates were so narrow as to exclude the possibility of a true case-control difference of any importance. Why do you believe this null result does *not* detract from the hypothesis that maternal folic acid deficiency is a cause of neural defects in her offspring?

Answer 2.10 It is likely that above a certain threshold of folic acid intake—one effectively reached by all American women after 1997—there is no further reduction in risk of neural tube defects in relation to maternal folic acid intake or blood levels. The inverse association between maternal intake (or levels) and neural tube defects already has been well demonstrated in women below this threshold; the 1998–2003 study simply is unable to address this question in the setting of population-wide high folic acid intake. The cases in the 1998–2003 study no doubt arose via a causal pathway not involving folic acid.

Question 2.11 In an effort to gauge the impact of surgery and radiation therapy on mortality among men with organ-confined prostate cancer that was well or moderately differentiated, a large cohort study was conducted.[7] Beginning one year after diagnosis, all-cause mortality in 65- to 80-year-old American men with this disease who received either surgery or radiation therapy (n = 32,022) was 69% that of the 12,608 men who received no definitive treatment, adjusted for tumor and demographic characteristics and for the presence of comorbidity (as assessed in Medicare claims data). For mortality from prostate cancer itself, the corresponding relative risk associated with receipt of active treatment was 0.67.

a. Based on the adjusted relative risk for all-cause mortality of 0.69 associated with receipt of active treatment for prostate cancer, and assuming that this reduced risk was indeed a result of treatment, how many of the 4,663 deaths in the untreated men in this study's follow-up period might have been prevented had they also been treated?
b. Perform the same calculation for the 314 deaths from prostate cancer itself, assuming a relative risk associated with receipt of treatment of 0.67.
c. Assume that classification of cause of death in men with prostate cancer is 100% accurate, and that in truth treatment of prostate cancer does *not* have an impact on mortality from other causes. What do you believe to be the most likely explanation for the difference in the numbers obtained in (a) and (b) above? Why?

Answer 2.11

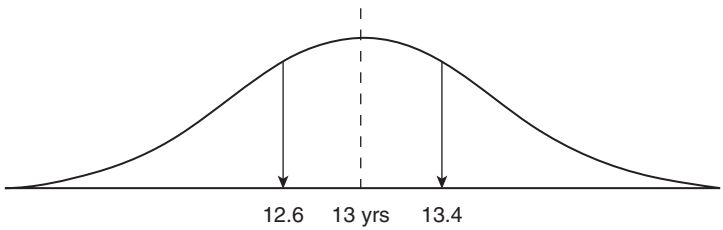
a. Deaths (all causes) in "observation" group = 4,663
   % potentially averted = 100% – 69% = 31%
   4,663 x .31 = 1,446
b. Deaths from prostate cancer in "observation" group = 314
   % potentially averted = 100% – 67% = 33%
   314 x .33 = 104
c. There were 1,446 – 104 = 1,342 fewer deaths from causes other than prostate cancer in men in the observational group than would have been predicted by the rates in the treatment group. The difference between the observed and expected number must be the result of confounding. Despite the considerable efforts of the authors to nullify confounding (e.g., by eliminating the first year of follow-up postdiagnosis, and by adjusting for demographic characteristics, disease characteristics, and the presence of comorbidity), an inherent mortality disadvantage must have been present in the untreated men.

Question 2.12 The following is an excerpt from an article in the JAMA:

> Age-at-menarche of 38 college female athletes was ascertained and related to the age of initiating training. The 18 premenarche-trained athletes had a mean menarcheal age of 15.1 ± 0.5 years, whereas the 20 postmenarche-trained athletes had a mean menarcheal age of 12.8 ± 0.2 years. Thus, it is apparent that strenuous exercise can delay the onset of menses to a substantial degree.

The authors' choice of comparison is likely to have produced a biased result. Why, and in what direction?

Answer 2.12 The bias results from the fact that the effect being measured—difference in age at menarche—is intertwined with the definition of the groups being compared—pre-versus postmenarche initiation of training. Thus, if a group of girls with age-at-menarche distributed normally about a mean of 13 years all began to train at 13, the mean age at menarche for the postmenarche trained would be <13 years while the mean age at menarche for the premenarche trained would be >13 years. For example:



12.6 = Mean age at menarche for "postmenarche trained"

13.0 = Age at which all started training

13.4 = Mean age at menarche for "premenarche trained"

This bias will exaggerate the influence of training in delaying age-at-menarche.

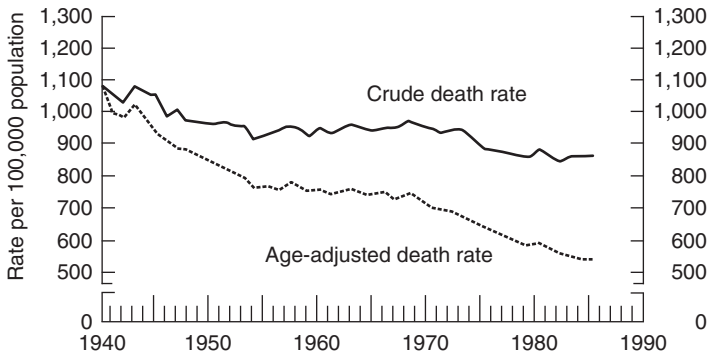
CHAPTER 3

Confounding

WHEN THE measured relation between an exposure and the occurrence of disease is distorted by the co-relation of each of these to another exposure or characteristic, we say that confounding is present. Some people, when they learn of an association based on data from one or more nonrandomized studies, do not consider the possibility of confounding, and automatically take the results at face value. Others cannot imagine that it is ever possible to disentangle the influence of one exposure from that of others with which it might be correlated, and so are unwilling to make any inferences from studies in humans that do not entail randomization. Certainly there are nonrandomized studies in which confounding is virtually absent, and others in which it is substantial and uncontrollable. However, most of the time we are in between these extremes, and it is possible to evaluate where confounding might be coming from, gauge its likely magnitude, and even explicitly take it into account in the study design and/or analysis. The questions in this chapter seek to illustrate how all of this can be done in specific circumstances, and also to illustrate more fully just what are the properties of a confounding variable.

This page intentionally left blank

Question 3.1 The following questions concern the data presented in the figure below:



Crude and age-adjusted death rates: United States, 1940-85

a. What is the reason for the increasing disparity between crude and age-adjusted death rates?
b. For purposes of age adjustment, can you tell which population distribution was used as a standard? Why, or why not?

Answer 3.1

a) The sharp fall in the age-adjusted death rate must reflect a fall in the age-specific rates on which it is based. The fact that the crude death rate fell to a much lesser degree must be a result of a shift in the age distribution over time, with groups at higher risk of death (generally the older segments of the population) being more heavily represented in later years.
b) The age distribution of the 1940 population was chosen as the standard, because only in that year were the crude and age-adjusted rates identical.

Question 3.2 The following data are taken from an article on the occurrence of out-of-hospital cardiac arrest in New York City during April 2002-March 2003[8]:

			Incidence per 10,000 person-years	
Race/ethnicity	*Population**	*No. of cardiac arrests*	*Crude*	*Age-Adjusted*
Black	1,393,859	1,257	9.0	10.1
Hispanic	1,489,208	636	4.2	6.5
White	2,345,564	1,908	8.1	5.8
Other	829,378	252	3.0	4.8

*18 years and older.

Can you draw any conclusion regarding the difference in age distribution between white and nonwhite residents of New York City ages 18 years and above during 2002–2003? If yes, how would you characterize that difference? If no, why not? (In all demographic subgroups of the population, the incidence of cardiac arrest rises sharply with increasing age.)

Answer 3.2 The white population must be relatively older because once the age difference between groups is accounted for, their rate of out-of-hospital cardiac arrest falls. This can happen only if the white population's age distribution put them at a relatively higher risk than that of members of other segments of the population, in other words, if whites were older on average.

Question 3.3 The following table presents the incidence of hospitalized pneumonia in the United States among persons 65 years and above during two periods of time:

	1988–1990		*2000–2002*	
Age (years)	*Population size*	*Incidence per 1,000 person-years*	*Population size*	*Incidence per 1,000 person-years*
65–74	17,506,833	10	18,495,139	12
75–84	9,939,111	21	12,409,949	26
≥85	2,939,646	49	4,413,680	51

An increased incidence was noted in each of the three age groups during 2000–2002 relative to that in 1988–1990. Could confounding by age have distorted the time trends *within* each age group to some degree? If yes, why, and in what manner. If no, why not?

Answer 3.3 Yes, residual confounding by age likely is present to at least some degree in these data, leading to an exaggeration of the true change over time:

a) Rates rise steeply with age.
b) Across the three age categories, there were proportionately more persons in the older group(s) in 2000–2002 than in 1988–1990. Thus, *within* each age group it is likely that the mean age was higher in 2000–2002 than it was earlier. And so, even if the actual rate of pneumonia had not risen at all, it would appear to have done so in each of the three 10-year age groups.

Question 3.4 In a survey of Americans 18 years and older conducted in 2004, 24.1% of white men and 23.9% of black men were current smokers of tobacco. Irrespective of race, the prevalence of tobacco smoking differed little by age, except for a sharp drop in men 65 years and older.

The small difference in current smoking (0.2%) between white and black men noted above does not take into account the fact that in 2004 the average age of white men was somewhat higher. After age adjustment, would you expect the interracial difference to be larger, smaller (or reversed), or unchanged? Why?

Answer 3.4 The adjusted difference will be larger. A greater proportion of white than black American men are in the ≥65-year age group, and these older men have the lowest prevalence of smoking. Removing the age disparity by means of adjustment will allow the higher prevalence of smoking of white men to become more evident.

Question 3.5 The following data come from a Swedish study of children born following in vitro fertilization during 1982–1995.[9] Records of all 26 childhood disability centers in Sweden were reviewed to identify children in this cohort who received services from these centers. In addition, a comparison cohort of children was selected from the entire Swedish Medical Birth Registry so as to be similar to the IVF cohort with respect to sex, year of birth, and hospital of birth.

	Disability	
	Yes	*No*
All Children		
IVF	101	5,579
Comparison cohort	119	11,241
Singletons only		
IVF	45	3,183
Comparison cohort	115	10,955
Nonsingletons		
IVF	56	2,396
Comparison cohort	4	286

a. What is the cumulative incidence of disability in children conceived by means of IVF relative to that in other children, both adjusted and not adjusted for singleton/nonsingleton birth?
b. Which of the above relative risks addresses the following questions:
   i. What is the impact of IVF on the incidence of disability?
   ii. What is the impact of IVF on the incidence of disability, beyond the impact of IVF to lead to multiple births?

Answer 3.5

a) Crude relative incidence $= \dfrac{101/5{,}680}{119/11{,}360} = 1.70$

Adjusted relative incidence =

$$\frac{(45 \times 11{,}070 \div 14{,}298) + (56 \times 290 \div 2{,}742)}{(115 \times 3{,}228 \div 14{,}298) + (4 \times 2{,}452 \div 2{,}742)} = 1.38^{*}$$

b) i. The crude relative incidence. Since multiple pregnancy is a consequence of IVF, it is not appropriate to treat multiple pregnancy as a confounding variable.

ii. The adjusted relative incidence. This tells us how much of an association would remain if, hypothetically, the excess risk of multiple pregnancy associated with IVF could be eliminated.

* Method of Mantel and Haenszel.[10]

Question 3.6 A randomized trial was conducted[11] in which 2,763 women with a history of coronary heart disease were randomly assigned in equal proportion to:

a. a regimen of 0.625 mg/day of conjugated estrogens and 0.25 mg/day of medroxyprogesterone acetate; or
b. a placebo.

> During an average follow-up of 4.1 years the incidence of the combined endpoint, myocardial infarction and death from coronary heart disease was identical in the two groups (relative risk = 0.99, 95% CI = 0.80–1.22).

A critic of the study contended that the results may have been confounded, since diagnostic tests to ascertain the extent of coronary and other vascular disease at the outset of this trial were not performed on the study participants. Do you share this concern? If yes, why? If no, why not?

Answer 3.6 Given that this is a randomized trial of 2,763 persons, the treatment and placebo groups would be expected to be extremely similar with respect to their distribution of severity of vascular disease. Failure to have documented that severity in study participants, and thus the failure to adjust for it, should have introduced no bias at all.

Question 3.7 Investigators at the CDC sought to determine the efficacy of the drug zidovudine (AZT) in preventing HIV infection in health-care workers who had sustained a percutaneous exposure (e.g., a needle stick) while caring for patients with AIDS.[12] Among health-care workers with a history of such an exposure, 27 persons who developed HIV infection and a sample of those who remained uninfected were compared with regard to their receipt of AZT soon after the incident.

The analysis suggested that four characteristics of the exposure, or of the patient being cared for, were particularly more common among cases than controls: (a) "Deep" injury, (b) Visible blood on the device that caused the wound in the health-care worker, (c) Injury during a procedure involving a needle in an artery or vein, and (d) Terminal illness in the source patient.

The results of the study pertaining to the receipt of AZT following percutaneous exposure are summarized in the following table:

Number of the other four risk factors	*Cases*		*Controls*	
	AZT	*No AZT*	*AZT*	*No AZT*
0	0	0	40	88
1	0	3	51	73
2	2	9	33	22
3, 4	6	7	7	6

Among health-care workers with percutaneous exposure to HIV, estimate the risk of HIV infection in persons who received postexposure AZT relative to the risk in those who did not: a) Adjusted for the number of other risk factors for HIV infection and b) Not adjusted for the other risk factors.

Why do these two estimates differ from one another?

Answer 3.7 From this case-control study, the relative risk associated with AZT can be estimated by the odds ratio (OR).

The adjusted OR (calculated here by the method of Mantel and Haenszel) considers only persons with at least one other risk factor, because no case failed to have at least one.

$$\text{Adjusted OR} = \frac{\frac{0\times73}{127}+\frac{2\times22}{66}+\frac{6\times6}{26}}{\frac{3\times51}{127}+\frac{9\times33}{66}+\frac{7\times7}{26}} = 0.27$$

$$\text{Crude OR} = \frac{8}{19} \div \frac{131}{189} = 0.61$$

The adjusted OR is well below the null, suggesting a beneficial influence of AZT in blocking transmission of HIV infection. The crude OR is not as low as the adjusted OR, because of the tendency (see the data in the table) for exposed health-care personnel with *other* risk factors for contracting HIV to have received postexposure AZT.

Question 3.8 In white married U.S. males, the annual incidence of prostate cancer is about 1 per 100,000 at ages 35 to 44 years, and about 100 per 100,000 at ages 55 to 64 years.

The following table presents the incidence of cancer of the prostate in white U.S. men, ages 45 to 54 years, in relation to marital status:

Marital status	*Rate per 100,000 per year*
Married at present	12.5
Never married	11.2
Widowed	20.7

Disregarding the possible role of chance, what do you believe to be the most likely *non*causal explanation for the observed high rate of prostate cancer among widowed men?

Answer 3.8 Because the incidence of prostate cancer rises dramatically with age (1 per 100,000 at 35–44 years and 100 per 100,000 at 55–64 years), the incidence rate is not likely to be uniform within the 45- to 54-year age category. It may be expected that the widowed men within this age category are older on average than the married at present or never-married men, thus accounting for their higher incidence rate. There is residual confounding by age.

Question 3.9 A population-based case-control study of bladder cancer was conducted in a part of the western United States in which some areas have arsenic levels in drinking water that are relatively high (about 100 micrograms per liter).[13] Persons diagnosed with this condition during 1994–2000 were identified through the records of cancer registries serving a 6-county area. Controls under age 65 years were recruited through random digital dialing of phone numbers; controls 65 years and over were obtained through the records of the Health Care Financing Administration. Information concerning a history of living or working in the areas in which arsenic levels were elevated was obtained by means of interviews with the study participants, as was information on potential confounding factors.

On average, income levels in cases were lower than those of controls.

In discussing their findings, the authors acknowledged this imbalance, and suggested it was "likely related to the increased participation rates among [potential controls] in higher socioeconomic brackets." Nonetheless, they concluded, "Other studies have shown little or no association between socioeconomic status and bladder cancer, suggesting that this variable is not likely to act as a substantial confounder."

In this study, do you believe it would be possible for "annual income" to be a confounding variable when assessing the potential association between arsenic in drinking water and the incidence of bladder cancer? If yes, under what circumstance? If no, why not? (Assume that income level has been measured without error.)

Answer 3.9 Income will be a confounder if it is related, positively or negatively, to ingestion of arsenic. Income is already related to case/control status, probably by virtue of differences in level of participation across socioeconomic strata (as the authors suggest). The fact that income level is not truly a risk factor for bladder cancer in the population at large is irrelevant, given the case/control disparity in the distribution of income in *this* study.

Question 3.10 The following table contains data from a New Zealand study in which automobiles involved in crashes and a random sample of other automobiles were compared for color.[14]

Association of car color with car crash injury in Auckland

Car color	*No. (%) of cases (n = 567)*	*No. (%) of controls (n = 588)*	*Univariate odds ratio*	*Multivariable odds ratio**
White	145 (25.6)	146 (25.9)	1	1
Yellow	31 (5.5)	15 (2.8)	2.0 (1.0 to 4.0)	0.8 (0.3 to 2.3)
Grey	52 (9.2)	61 (10.0)	0.9 (0.6 to 1.5)	0.6 (0.3 to 1.3)
Black	36 (6.4)	34 (5.5)	1.2 (0.7 to 2.0)	2.0 (1.0 to 4.2)
Blue	91 (16.1)	96 (17.4)	0.9 (0.6 to 1.4)	0.9 (0.5 to 1.6)
Red	85 (15.0)	82 (13.3)	1.1 (0.7 to 1.8)	0.7 (0.4 to 1.4)
Green	42 (7.4)	44 (7.0)	1.1 (0.6 to 1.8)	1.8 (1.0 to 3.6)
Brown	55 (9.7)	49 (6.8)	1.4 (0.8 to 2.5)	2.1 (1.1 to 4.2)
Silver	30 (5.3)	61 (11.3)	0.5 (0.3 to 0.8)	0.4 (0.2 to 0.9)

*Adjusted for driver's age, ethnicity, alcohol consumption in past 6 hours, seat belt use, vehicle speed, average driving time each week, driving license status, vehicle insurance status, and weather.

From the information contained in the table, which one of the following statements do you believe to be true? Explain your answer.

a) On the basis of their age, ethnicity, alcohol consumption, and so forth (see footnote to the table), drivers of black or brown cars in Auckland during 1998–1999 tended to be at *higher* risk of a car crash injury than drivers of white cars.
b) On the basis of their age, ethnicity, alcohol consumption, and so forth (see footnote to the table), drivers of black or brown cars in Auckland during 1998–1999 tended to be at lower risk of a car crash injury than drivers of white cars.
c) From the data presented, no inferences can be made from these data regarding the underlying differences in risk between drivers of black/brown versus white cars.

Answer 3.10 The correct answer is (b). Since adjustment for the characteristics listed in the footnote led to an increase in the OR associated with driving a black or brown car, the underlying risk of a car crash injury must have been low relative to drivers of white cars.

Question 3.11 The following table is from an article that describes the prevalence at birth of a particular congenital malformation in Ontario, Canada, before and after a program of folate fortification of cereal grain products was begun in January, 1998[15]:

	Before fortification	*After fortification*	*Crude prevalence ratio (95% CI)*	*Age-adjusted prevalence ratio (95% CI)*
Length of observation (months)	48	29		
Maternal age (mean, SD) (years)	30.1 (0.16)	30.9 (0.081)		
Number of women	218,977	117,986		
Number of women with an affected child	248	69		
Prevalence (per 1,000 infants)	1.13	0.58	0.52 (0.40–0.67)	0.62 (0.46–0.83)

From the data presented in the table, what can be concluded regarding an association between maternal age and the occurrence of this malformation? Explain your answer.

a) On average, the risk rises with increasing maternal age.
b) On average, the risk falls with increasing maternal age.
c) The risk is unaffected by maternal age.
d) No conclusion can be drawn about a relation of maternal age to the prevalence of the malformation from these data.

Answer 3.11 The crude and age-adjusted relative risks associated with being born after the introduction of fortification differ, so maternal age must be related to the occurrence of the malformation. And, since we know that:

a) women in the latter time period were, on average, older than those in the earlier period;

and

b) adjustment for maternal age made the prevalence ratio rise,

older women must be a lower risk age group. The correct answer is (b).

Question 3.12 The following is an excerpt from an article, "Cancer beats a retreat," that appeared in a 1998 issue of *US News and World Report*. It dealt with recent declines in cancer incidence rates in the United States.

> The researchers acknowledge that the cancer war might not be doing as well as the new data suggest. They adjusted their numbers on the basis of the age of the U.S. population in 1970, a widely used standard in cancer studies. However, the population has aged in the past 25 years. Since cancer disproportionately strikes older people, using a current standard might yield less encouraging results.

Is it possible that using the age distribution of the U.S. population in 1990 as a standard "might yield less encouraging results" than using the 1970 population? If so, how could this occur? If not, why not?

Answer 3.12 The use of the 1990 U.S. population as a standard (which gives more weight to the rates in older age groups) would tend to give a relatively smaller estimate of the declining rate only if the size of the decline were smaller in older than in younger persons.

Question 3.13 The following is excerpted from a letter to the editor of a medical journal:

> There is a sharp contrast in the consistency of success in studies that have sought genotype-phenotype associations in animals and in humans. For example, animal models of depression and anxiety disorders have consistently demonstrated genotype-phenotype associations. By contrast, a recent genome-wide association study (GWAS) of depression found no significant associations. One central difference between these 2 research approaches lies in control over potentially relevant environmental exposures. These exposures are effectively randomized in animal models, but such control is absent from observational human gene-hunting studies.

Assume that the genome-wide association studies were conducted in racially homogeneous populations. Do you agree that the lack of randomization in these studies is likely to be responsible for the difference between their results and those obtained in studies conducted in other species? If yes, why? If no, why not?

Answer 3.13 No. Whatever environmental exposures are related to the development of depression and anxiety disorders in human beings, these almost certainly do not bear on a person's genotype. Given the likely absence of confounding, the lack of randomization of genotype in human studies of depression should not affect the validity of the results obtained.

Question 3.14 The data presented in the following table come from in-person interviews of random samples of the U.S. population ages 19 years and above. They indicate that the prevalence of cigarette smoking in 2008 was greater among American Indian/Alaska Native persons than among non-Hispanic whites. The data are not adjusted for age, however, and the proportion of American Indians/Alaska Natives over 64 years of age in those years was smaller than the corresponding proportion of non-Hispanic whites. If age adjustment were to be done, would you expect the difference in smoking prevalence between these two demographic subgroups to increase, shrink, or remain the same? Explain.

Percentage of persons aged ≥18 years who were current cigarette smokers, by sex,race/ethnicity, and age — National Health Interview Survey, United States 2008

	Men (n = 9,387)	*Women (n = 12,138)*
Characteristic	% (95% CI)	% (95% CI)
Race/Ethnicity		
White, non-Hispanic	23.5 (22.2–24.9)	20.6 (19.3–21.9)
American Indian/ Alaska Native	42.3 (27.4–57.2)	22.4 (12.5–32.3)
Age (yrs)		
18-24	23.7 (20.3–27.1)	19.0 (16.2–21.8)
25–44	26.4 (24.5–28.2)	21.1 (19.5–22.7)
45–64	24.8 (22.8–26.7)	20.5 (18.9–22.1)
≥65	10.6 (8.8–12.3)	8.4 (7.1–9.6)

Answer 3.14 The difference would be expected to shrink. Younger Americans tend to be current smokers more than older ones (see table), and American Indians/Alaska Natives tend to be younger than whites. Thus, a part of the difference in smoking prevalence is attributable to age alone, and that part would disappear with age adjustment.

Question 3.15 The following appeared in the November 13, 1999, issue of *Lancet*:

> The general health of the East German population was below the standard of their Western counterparts despite an emphasis on public health. In 1991, life expectancy was 3.2 years shorter in men (69.9 years) and 2.3 years in women (77.2 years) than in West Germans. Mortality in 1991 also showed important differences: 1324 per 100,000 population in East German women, 1215 in men compared with 1149 in West German women and 1061 in men.

In trying to determine whether the risk of death differed between East and West Germans in 1991, you'd be loath to use the mortality rates provided in this article. Why do you believe they might be misleading?

Answer 3.15 The mortality rates for 1991, higher in women than in men in both populations, must not be age-adjusted. (We know this because the greater life expectancy in German women than German men must be a result of the age-specific mortality among women being *low* relative to that of men.) Therefore, if the East German population had been older, on average, than the West German population, some or all of the difference in the crude mortality rates could be due to confounding by age.

Question 3.16 I came across the following statement (and data) in a draft of a master's thesis a *long* time ago. Can you think of a reason that the observed association between marital status and suicide is almost certainly greatly overestimated?

> In general, suicide is less frequent among the married—except for the young married population. As the table indicates, in the under-20 age group, suicide rates are higher in married or divorced persons than in single persons. Explanations for this phenomenon may be a desire to escape unsatisfactory home conditions or pregnancy resulting in an unplanned, unhappy marriage.

Death rates from suicide by marital status and sex, persons less than 20 years of age: United States, 1949–1951

	Total	*Single*	*Married*	*Divorced*
White males	0.9	0.9	6.2	14.5
White females	0.4	0.3	3.1	13.8

Answer 3.16 There is an unusually great degree of confounding by age here. Persons at the upper end of the 0- to 20-year age groups are the ones relatively most likely to get married (and divorced) and (irrespective of marital status) to commit suicide.

Question 3.17 The following data describe the occurrence of primary cesarean delivery (i.e., in women with no prior cesarean delivery), expressed as a percentage of all live births, in the United States in 1990:

	Age of mother						
	Total	*Under 20 years*	*20–24 years*	*25–29 years*	*30–34 years*	*35–39 years*	*40–49 years*
All births	16.0	14.7	15.0	16.0	16.5	19.0	23.5
1st	24.6	16.9	22.6	27.1	32.2	39.2	46.9
2nd	8.9	6.8	7.3	8.7	10.4	13.6	20.1
3rd	8.4	6.3	6.5	7.7	9.0	12.0	18.1
4th and over	8.8	7.6	6.8	7.4	8.8	11.2	15.1

There is generally a 2- to 3-fold difference in the proportion of cesarean deliveries between the youngest and the oldest categories of maternal age within individual categories of live birth order. However, the difference is considerably smaller—14.7 percent versus 23.5 percent—when the comparison is made for all birth orders combined.

What do you believe to be the explanation for the relative difference in the frequency of cesarean delivery across maternal age being so much smaller when examined overall than within individual birth order categories? Why?

Answer 3.17 There is confounding by birth order: (a) at any maternal age, first births are associated with relatively high percentage of cesarean deliveries; (b) young mothers, relative to older ones, would be expected to be in the lower birth orders, and on that basis alone would be at increased risk of cesarean section. The confounding attenuates to some degree the very strong association seen in the birth-order specific results.

Question 3.18 In a case-control study that was based on information obtained from spouses, a history of high-intensity leisure-time physical activity (LTPA) during the prior year was associated with a reduced risk of primary cardiac arrest.[16] Relative to the risk in persons with light or no LTPA, that in persons with high-intensity activity was 0.19; after adjusting for age, smoking, education, diabetes, hypertension and self-reported health status ("fair," "good," "excellent," or "very good"), the corresponding relative risk was 0.36.

There was a case-control difference in self-reported health status, and it is reasonable to assume that adjustment for this variable explained at least some of the difference between the unadjusted and adjusted odds ratios associated with high-intensity LTPA. However, health status is difficult to assess from interview data alone, so some subjects were probably misclassified on this variable. Had no misclassification been present, what would you predict the adjusted odds ratio associated with high intensity LTPA to be?

a. 0.36
b. Greater than 0.36
c. Less than 0.36
d. No prediction is possible.

Explain your answer.

Answer 3.18 Inaccurate measurement of a confounding variable will give rise to an adjusted risk estimate that is spuriously close to the unadjusted one. So, more complete adjustment for self-reported health status would be expected to lead to a relative risk associated with high LTPA that is greater than 0.36.

Question 3.19 Investigators in Australia surveyed 202 pregnant women who were undergoing abortion who were using oral contraceptives (OCs) when they became pregnant. Among other things, the investigators gathered information on type of OC used by these 202 women, in comparison with the types of OCs used by all OC users in Australia. They were specifically interested in pregnancies among women using triphasic OCs. They presented the following results:

	Abortion-seeking OC users	*All Australian OC users*
Mean age	*23 years*	*30 years*
Type OC used		
Triphasic	52%	42%
Monophasic		
30 microgram ethinylestradiol	27.2%	29.2%
50 microgram ethinylestradiol	7.9%	13.2%
Norethisterone	9.4%	9.0%
Progestin-only	3.5%	6.2%

The investigators found that triphasic OC use was more common ($P < 0.01$) in OC users seeking abortion than would be expected based on the national usage. They concluded that triphasic OCs were more likely than other OCs to increase the risk of inadvertent pregnancy (and they hypothesized this was because they have a smaller margin of safety due to decreased progestin content).

You are concerned that the age difference between the two groups above could be distorting the results. Under what circumstance would this be so?

Answer 3.19 These data indicate that OC users seeking abortion are younger than OC users overall. If young age is associated with an increased risk of abortion (either because of a greater risk of pregnancy, or among pregnant women, a greater use of induced abortion) and if younger OC users tend to use triphasic preparations relatively more than older women do, it is possible that there is no true association between triphasic OC use and inadvertent pregnancy. Information on the type of OC use, by age, is necessary to determine if this is the case.

Question 3.20 A case-control study of the relationship between asthma and a history of pertussis among children (2 years of age) was carried out in one community. Although overall participation was good, not all parents of cases and controls could be interviewed. A difference in the percentage interviewed based on case/control status and on day care enrollment was noted. Only 63% of control parents with a child enrolled in day care could be interviewed, versus 82% of control parents with a child who was not enrolled in day care. In all, 93% of case parents were interviewed, a percentage that was the same for parents of children who were and were not enrolled in day care.

In this study, under what circumstance would "enrollment in day care" confound the association between asthma and pertussis vaccination?

Answer 3.20 Because there was differential case-control response by day care enrollment status, day care enrollment would be a confounder if it were also associated with pertussis vaccination status.

Question 3.21 A case-control study of breast cancer in relation to prior use of postmenopausal hormones was conducted among American women enrolled in a network of health insurance plans.[17] Women who were newly diagnosed with breast cancer and who had been enrolled for at least 2 years prior to that date comprised the case group.

For each case, four women were selected as controls, matched on year of birth and enrollment status as of the time of the case's diagnosis (and for the 2 years prior to that time).

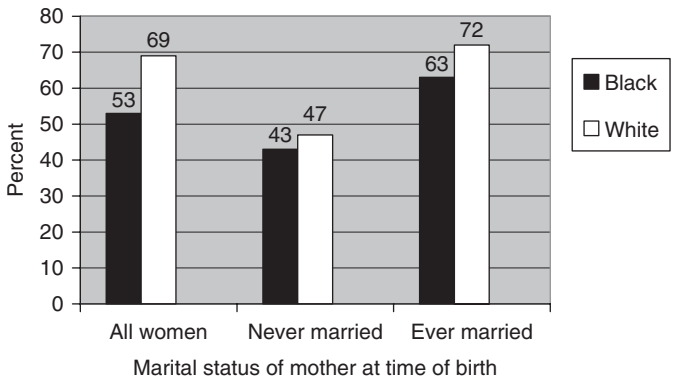
Data on hormone use were obtained from paid claims for pharmaceuticals. While a similar proportion of cases and controls had taken estrogen alone prior to the time of the cases' diagnoses (odds ratio = 0.96, referent category being hormone nonusers), a higher portion of cases than controls had taken combined estrogen-progestin therapy (odds ratio = 1.44, referent category being hormone nonusers).

The information available on study subjects did not include whether they had a history of bilateral oophorectomy. Such a history is associated with a reduced risk of breast cancer. If a woman has not undergone bilateral oophorectomy and is prescribed hormone therapy, generally that therapy will be in the form of a combined estrogen-progestin regimen. Women whose ovaries have been removed generally will be given estrogen alone.

Is there reason to believe that the two odds ratios obtained are biased due to the investigators' inability to adjust for a history of bilateral oophorectomy? If yes, why, and in which direction? If no, why not?

Answer 3.21 The inability to adjust for a history of bilateral oophorectomy will lead to bias (confounding), at least to some degree. Based on their absence of ovarian tissue, the users of estrogen alone are a low-risk group for breast cancer: therefore, the odds ratio associated with use of estrogen alone that was obtained in this study is spuriously low. In contrast, users of combined therapy have an inherently elevated risk, since they do have intact ovaries. As a result, this study's odds ratio associated with receipt of hormone therapy is likely to be falsely high.

Question 3.22 Based on the figure below, which one of the following statements is correct for American women giving birth during 1979–1982:



Percentage of mothers receiving prenatal care in the first trimester, by race and marital status: United States, 1979-1982

a) A greater proportion of white than black mothers had never been married.
b) A greater proportion of black than white mothers had never been married.
c) An equal proportion of black and white mothers had never been married.
d) From the data provided, no conclusion as to the marital status distribution of white and black mothers is possible.

Explain your answer.

Answer 3.22 The correct answer is (b). Overall, 69% of white women received prenatal care, a weighted average of the figures of 47% for never-married women and 72% for ever-married women. The preponderance of these women must have been ever-married,* to explain the fact that the overall percentage and never-married percentage are so similar. In contrast, only about half the black women had ever married: the percentage receiving prenatal care overall—53%—lies halfway between the percentages for never- and ever-married women.

* x = proportion of white women who were ever married

$72x + 47(1-x) = 69$

$x = 0.88$

Question 3.23 The following data come from a 1975 survey of a representative sample of about 140,000 persons from the United States population:

	% cigarette smokers	
Age (years)	*Male*	*Female*
21–24	41.3	34.0
25–34	43.9	35.4
35–44	47.1	36.4
45–54	41.1	32.8
55–64	33.7	25.9
65–74	24.2	7.1

	% cigarette smokers	
Marital Status	*Male*	*Female*
Married	38.3	28.3
Single	37.5	30.6
Divorced or separated	60.1	50.0
Widowed	35.7	19.3

From these data, can you conclude that there is an association, beyond that which could be explained on the basis of age: (a) between smoking and widowhood among females? (b) among persons of both sexes, between smoking and being divorced or separated? (No calculations are required.)

Answer 3.23

a) Widowed women are likely to be considerably older, on average, than other women, and (at the time the data were collected) older women were less often cigarette smokers than younger women. Thus, it is possible that after controlling for age, widows may not have had a smaller proportion of cigarette smokers than women in the other marital status categories.
b) The frequency of cigarette smoking among divorced/separated persons was high—50%–60%—relative to that of any age groups in the population as a whole. No matter what the age distribution of divorced/separated persons, a prevalence of cigarette smoking of 50%–60% cannot be explained solely by of an overrepresentation of a particular age group. An association beyond that which could be explained by age must have been present.

Question 3.24 This question is based on the following (slightly paraphrased) abstract[18]:

> *Background.* Among women pregnant for the second time, the risk of pre-eclampsia is lower than in their first pregnancy, but not if the mother has a new partner for the second pregnancy. One explanation is that the risk is reduced with repeated maternal exposure and adaptation to specific antigens from the same partner. However, the difference in risk might instead be explained by the interval between births.
>
> *Methods.* We used data from the Medical Birth Registry of Norway, a population-based registry that includes births that occurred between 1967 and 1998. We studied 551,478 women who had two or more singleton deliveries.
>
> *Results.* The risk of pre-eclampsia in a second pregnancy was directly related to the time that had elapsed since the preceding delivery, and when the interbirth interval was 10 years or more, the risk approximated that among women pregnant for the first time. In unadjusted analyses, a second pregnancy involving a new partner was associated with higher risk of pre-eclampsia than a second pregnancy with the same partner, but after adjustment for the interbirth interval, the difference in risk of pre-eclampsia was reduced.

From the foregoing, what can be concluded about the relation between interbirth interval and the presence of a new partner? Explain your answer.

a) Women with new partners tended to have a shorter interbirth interval than women with the same partner.
b) Women with new partners tended to have a relatively longer interbirth interval.
c) Women with or without new partners had, on average, the same interbirth interval.
d) Nothing can be concluded about a possible association between interbirth interval and the presence of a new partner.

Answer 3.24 The correct answer is (b). The presence of a new partner must be associated with a longer interbirth interval. The crude elevated risk for pre-eclampsia present in women with new partners was diminished once adjustment was made for interbirth interval. Therefore, since long interbirth intervals are associated with high risk, these must have been relatively more common in women with new partners.

CHAPTER 4

Cohort Studies

COHORT STUDIES compare the occurrence of an illness or injury between persons with and without an exposure or characteristic. Threats to the validity of cohort studies can come from inaccurate characterization of exposure status; incomplete or inaccurate ascertainment of health outcomes; and differences between exposed and unexposed persons with respect to factors that themselves bear on the occurrence of the outcome. The forward-looking structure of cohort studies—start with exposure, follow subjects for the development of an outcome event—resembles that of randomized trials, and that similarity can put you off your guard when trying to interpret the results of such studies. The questions that follow are intended to put you back on guard.

This page intentionally left blank

Question 4.1 In a study of more than 20,000 women who had received a cosmetic breast implant (one that was not provided in relation to breast cancer surgery), all-cause mortality beginning 1 year later was only 74% that of socioeconomically and demographically comparable women in the population as a whole (95% CI = 0.68–0.81).[19] However, all-cause mortality in the implant recipients was nearly identical to that of some 16,000 women beginning 1 year after having undergone another form of cosmetic surgery during the same period of time (relative risk = 1.02, 95% CI = 0.89–1.17, adjusted for demographic characteristics).

Despite the observed 26% reduction in mortality compared to women in general, the authors concluded that the receipt of breast implants does not appear to influence death rates. Do you agree? If yes, why? If no, why not?

Answer 4.1 Women who have a serious illness are unlikely candidates for cosmetic implants; the proportion of women with such an illness is undoubtedly lower in the implant cohort than in the population as a whole (even after restricting the period of observation by excluding the first year after the operation). Thus, even in the absence of any influence of breast implants on the risk of death, the mortality experience of women receiving them would be expected to be more favorable than that of women in general. A better basis for comparison would be the death rates for the women undergoing other forms of cosmetic surgery, in which the same selection factors would be expected to be present. The similarity of death rates between implant recipients and other women undergoing cosmetic surgery argues that receipt of breast implants has no bearing on the risk of death.

Question 4.2 A study of the efficacy of pneumococcal vaccination in the elderly was described as follows[20]:

> We conducted a 2-year retrospective cohort study among all elderly members of a staff-model managed care organization who had a baseline diagnosis of chronic lung disease. The study outcomes were assessed over 2 years, from November 15, 1993, through November 14, 1995, and included hospitalizations for pneumonia and influenza.
>
> Of 1898 subjects, 1280 (67%) had received pneumococcal vaccination. This included 843 (44%) who were vaccinated prior to November 15, 1993 and an additional 437 (23%) vaccinated after that date. During the follow-up period there were 174 hospitalizations for pneumonia and influenza. The observed cumulative incidence was 138 per 1000 in the 618 unvaccinated persons and 70 per 1000 in the 1280 persons following their receipt of vaccination. This represents a 49% reduction in hospitalization for pneumonia and influenza.

Even if there were neither misclassification nor confounding in this study, the estimate of benefit associated with pneumococcal vaccination in elderly patients with chronic lung disease must be biased.

a. Why? In which direction?
b. How could the analysis be conducted to remove this source of bias?

Answer 4.2

a. Not all patients in this study were followed for the same length of time. Specifically, the 437 patients vaccinated during the 2-year study period were at risk for hospitalization only for the period after immunization. Thus, even had the hospitalization *rates* been the same in vaccinated and unvaccinated groups, the *cumulative incidence* would be lower in the former group. This would lead to an overestimate of the vaccine's efficacy.
b. The analysis should employ a person-time denominator, so that incidence rates can be calculated. The 437 persons vaccinated after November 15, 1993, would contribute person-time-at-risk to the experience of the *un*vaccinated persons until their date of vaccination, and to the experience of vaccinated persons afterward. Because the risk of hospitalization for pneumonia and influenza may be relatively low in the several week period immediately after November 15 (i.e., before the seasonal peak), adjustment for calendar period may be needed. Otherwise, confounding could arise from the different calendar distribution of person-time between vaccinated and unvaccinated groups.

Question 4.3 The following is taken from the abstract of an article on the mortality experience of employees at a polymer manufacturing facility, the DuPont Washington Works plant in West Virginia[21]:

> *Methods*. The cohort comprised 6,027 men and women who had worked at the facility between 1948 and 2002; these years delimit the mortality follow-up period. Standardized mortality ratios (SMRs) were estimated to compare the observed number of deaths to expected numbers derived from mortality rates for 2 reference populations: the West Virginia state population and an 8-state regional employee population from the same company.

The results of the study for deaths from heart disease are shown in the table below. (The SMRs presented are adjusted for age, sex, and calendar time.)

SMR estimates with 95% confidence intervals for mortality for heart disease for all Washington Works employees compared to 2 external reference populations

	WW cohort	*WV population*			*DuPont 8-state regional employee population*		
Cause of death	*O*	*E*	*SMR*	*95% CI*	*E*	*SMR*	*95% CI*
All heart disease	314	475.6	66.0	58.9, 73.7	284.5	110.4	98.5, 123.3

WW = Washington Works; WV = West Virginia; O = observed deaths; E = expected deaths; SMR = standardized mortality ratio = O/E x 100%; CI = confidence interval.

The confidence intervals around the SMRs based on expected deaths in the DuPont Region 1 workers are somewhat wider than those based on expected deaths in West Virginia. Nonetheless, when evaluating the possible impact of employment at the Washington Works plant on mortality from heart disease, why might the SMRs based on death rates in DuPont Region 1 workers provide a more valid estimate?

Answer 4.3 Persons with heart disease are less likely than other persons to become employed and stay employed, and also are more likely to die of heart disease. The proportion of WW employees with heart disease is almost certainly smaller than in the West Virginia population, leading to a spuriously low estimate of relative mortality from heart disease when these population death rates are used as a basis for comparison. A similar distortion would not be expected to be present using mortality rates of other workers as a means of determining the expected number of deaths.

Question 4.4 In Paris in the 1830s, bloodletting was believed to be efficacious in the treatment of pneumonia (and a number of other diseases). Virtually all patients diagnosed with pneumonia underwent this treatment, some early in the course of the disease and some later. Louis compared the case-fatality in 41 patients with pneumonia who were bled within the first 4 days of disease onset with that in 36 others who were bled later. The number of deaths were 18 and 9 in the two groups (case-fatality = 44% and 25%, respectively). While the early-bled cases were somewhat older than the late-bled cases (41 vs. 38 years), Louis concluded that "the effect of venesection [i.e., blood letting] on the progress of pneumonitis is much less than is commonly thought."

Presented with these data, today's epidemiologists would exclude from the analysis the experience during the first 4 days following disease onset, and would not calculate case-fatality from disease onset but rather mortality rates starting on day 5 (for the early-bled groups) or from the time bloodletting began (late-bled group), adjusting for number of days since disease onset. What would be the rationale for this modified approach? Do you expect that today's method would raise or lower the estimate of relative risk of death associated with receiving early blood letting?

Answer 4.4 A comparison of case-fatality from the time of disease onset inflates then-relative case-fatality associated with early bleeding, given that deaths that occur during the 4 days following disease onset can be only in the early-bled group. The rate-based approach that takes this into account, and also adjusts for time since disease onset (after four days), will produce a relative risk of death associated with early blood letting that is lower—as long as some deaths do occur within the first four days after illness onset.

Question 4.5 This question is based on the following abstract (abridged):

> *Methods* We conducted a retrospective cohort study of postwar mortality according to cause among 695,516 Gulf War veterans and 746,291 other veterans. The follow-up continued through September 1993. A stratified, multivariate analysis (with Cox proportional-hazards models) controlled for branch of service, type of unit, age, sex, and race in comparing the two groups. We used standardized mortality ratios to compare the groups of veterans with the general population of the United States.
>
> *Results* Among the Gulf War veterans, there was a small but significant excess of deaths as compared with the veterans who did not serve in the Persian Gulf (adjusted rate ratio, 1.09; 95 percent confidence interval, 1.01 to 1.16). In both groups of veterans the mortality rates were lower overall than those in the general population. The adjusted standardized mortality ratios were 0.44 (95 percent confidence interval, 0.42 to 0.47) for Gulf War veterans and 0.38 (0.36 to 0.40) for other veterans.
>
> (N ENGL J MED 1996;335:1498-504.)

The veterans who served in the Persian Gulf War had a postwar mortality that was 9% higher than that of other veterans, but only 44% that of Americans in general. Which of these figures do you believe is most likely to reflect the subsequent impact (if any) of having experienced the Persian Gulf War? Why?

Answer 4.5 A comparison of mortality in Gulf War veterans to that in similar-aged Americans in general would likely be biased, since persons with a number of diseases would not be eligible for military service and would also be at an increased risk of death during 1991–1993. The more appropriate comparison group is that of veterans who did not serve in the Persian Gulf (which produced the rate ratio of 1.09).

Question 4.6 A cohort study was conducted among 6,849 Swedish men with localized prostate cancer diagnosed during 1997–2002.[22] Death rates through 2008 were compared between those who did and did not undergo treatment with curative intent (most commonly, a radical prostatectomy). The cumulative 10-year mortality from prostate cancer was low (2.7 per 100) in the men receiving an attempt at cure, and only 0.9 per 100 higher than this in those not so treated. About 10 per 100 of the actively treated men died of causes other than prostate cancer; adjusting for age, the corresponding figure for the men in whom a curative procedure was not attempted was nearly twice as high.

The very large observed difference in mortality from causes other than prostate cancer was almost certainly not the result of chance, and it seems unlikely to be a result of an attempt to cure localized prostate cancer. What do you believe to be the explanation?

Answer 4.6 When interpreting the results of any cohort study, it is always necessary to ask, "Could some reason for the presence or level of exposure itself bear a relation to the outcome in question?" In this instance, it seems likely that one factor entering into the decision to attempt a curative procedure in a man with localized prostate cancer is the perceived life expectancy of that man: The presence of a life-shortening condition (e.g., heart disease, some other form of cancer) would argue against an intervention that (even though potentially curative) would entail its own morbidity. As a result of this source of confounding, the observed reduction in mortality from causes other than prostate cancer associated with active treatment almost certainly does not reflect a benefit of that treatment.

Question 4.7 This question is based on an excerpt of an abstract of a published article[23]:

> *Background*. Reports on the relation between anthropometric variables (height, weight) and physical activity with ovarian cancer have been inconclusive. The objective of the current study was to extend investigation of potential associations in the Iowa Women's Health Study cohort.
>
> *Methods*. The relation between self-reported anthropometric variables and incident ovarian cancer was studied in a prospective cohort of women ages 55–69 years who were followed for 15 years. Two hundred twenty-three incident cases of epithelial ovarian cancer were identified by linkage to a cancer registry.
>
> *Results*. No association was found overall between ovarian cancer and height. Although current body mass index (BMI) was not associated with ovarian cancer, a BMI ≥30 kg/m2 at age 18 years appeared to be associated positively with ovarian cancer (multivariate-adjusted RR, 1.83 for BMI ≥30 kg/m2 vs. BMI <25 kg/m2; 95% CI, 0.90–3.72), and this association was stronger after exclusion of the first 2 years of follow-up (RR, 2.15; 95% CI, 1.05–4.40).
>
> *Conclusions*. Anthropometric variables were not major risk factors for ovarian cancer in the cohort studied; however, high BMI in early adulthood may increase the risk of ovarian cancer among postmenopausal women.

In some cohort studies, events and person-time in exposed and unexposed persons begins only after a period of time has elapsed following formation of the cohort. When evaluating the possible influence of BMI at age 18 years as a predictor of risk of ovarian cancer in the Iowa Women's Health Study, would you recommend that approach, or instead one in which events and person-time are tabulated right away? Why?

Answer 4.7 The exclusion of events for a period of time following the formation of a cohort is appropriate if there is concern that there are occult cases of disease among cohort members in whom exposure status ascertained at the beginning of follow-up does not reflect that at the time their disease was being produced. For the variable BMI at age 18, this is of no concern, given that the earliest a woman's follow-up could begin was at age 55. The appropriate analysis in this study should consider events and person-time from the very start of follow-up.

Question 4.8 A British study enrolled a large number of postmenopausal women attending a breast cancer screening program and ascertained whether they were or were not taking hormone replacement therapy (HRT) at that time.[24] During the follow-up period, 2,894 users of HRT at baseline developed breast cancer, versus 3,202 never users, producing a relative risk of 1.66. Deaths from breast cancer occurred in 238 users of HRT at baseline and in 191 never users (relative risk = 1.22).

A letter to the editor (*Lancet* 2003;362:329) argued that

> the data provided for breast-cancer mortality are somewhat misleading. Compared with never-users of HRT the relative risk of death from breast cancer was raised in current-users. However, this finding should not be interpreted as evidence that HRT increases the risk of mortality among women diagnosed with breast cancer; the overall breast-cancer mortality rate will necessarily be higher in current-users as a result of the higher frequency of breast cancer among such women. To quote breast-cancer mortality figures for the subgroup of women diagnosed with breast cancer would seem more appropriate—i.e., never-users of HRT, mortality rate 8.2% (238 of 2894) versus current-users of HRT, mortality rate 6.0% (191 of 3202). These figures give a crude relative risk estimate of 0.725 for current-users versus never-users for breast-cancer mortality, indicating a lower risk of death in women taking HRT at the time of their diagnosis with breast cancer than in never-users.

Which of the two approaches to gauging mortality from breast cancer in women who use HRT—that of the authors of the study, or that of the authors of the letter to the editor—has the potential to provide data that are "somewhat misleading"? Why?

Answer 4.8 The data produced by the authors of the letter are misleading. A comparison of the risk of death from breast cancer in women who do and do not receive HRT needs to incorporate the possible influence of HRT on both incidence and on case-fatality. The use of mortality rates in users and nonusers by the authors of the article does this; the use of case-fatality alone does not.

Question 4.9 A meta-analysis of 13 studies of mortality from brain cancer among firefighters obtained a summary standardized mortality ratio (SMR) of 1.09 (95% CI = 0.92–1.25).[25] Using data from six of these studies that provided results based on duration of employment, a second meta-analysis conducted by the same author showed the following:

Duration of employment (years)	*Observed*	*Expected*	*SMR*
<10	8	6.50	1.23
10–19	12	7.41	1.62
20–29	11	6.30	1.75
≥30	11	5.24	2.10

Put aside the possible concern with "healthy worker" or other form of bias in these studies. Also, assume that both duration of firefighting employment and death from brain cancer have been ascertained without error. Explain why the SMRs above likely overstate the association between any particular duration of firefighting and mortality from brain cancer.

Answer 4.9 A weighted average of the SMRs in the table would give an overall SMR considerably greater than the value of 1.09 that is based on the results of all studies. The results of the six studies are not representative of the entire group of 13, most likely as a result of a form of publication bias—the tendency of authors reporting the results of individual studies to preferentially present positive results. Plausibly, the authors of one or more of the remaining seven studies examined SMRs by duration of employment, obtained a null or inverse relation, and chose to provide to readers only the data for all durations combined.

Question 4.10 In a randomized controlled trial of screening for two forms of cancer (breast and colon), more than 150,000 men and women were recruited to take part. Follow-up of these individuals took place over an average of 5 years for cancer incidence and cause-specific mortality.

In an analysis unrelated to those bearing on the efficacy of screening, incidence and mortality rates in all trial participants combined were compared to those of demographically comparable individuals in the population as a whole. The incidence of cancer (excluding cancer of the breast and colon) in the trial participants was 89% that of the general population, whereas the corresponding figure for cancer mortality was 56%.

Because of the large number of events, chance is a highly unlikely explanation for the difference between the relative risk for cancer incidence (0.89) and that for cancer mortality (0.56). What do you believe to be the most likely explanation for the difference?

Answer 4.10 Persons dying of cancer are unlikely to be participants in a cancer screening program. Their exclusion from the study participants, but not from the comparison population, would be expected to have a large impact on relative cancer mortality but little or none on relative cancer incidence. This represents a form of "healthy screenee bias."[26]

Question 4.11 The occurrence of transient neurologic symptoms (transient ischemic attacks, TIA) commonly heralds the later occurrence of a stroke (with its associated nontransient neurologic damage). In 1,707 patients with a TIA, a stroke took place in the subsequent 90 days in 14% of the 235 who received warfarin anticoagulation therapy at the time of the TIA and in 10% of the other TIA patients ($p = .04$). The latter patients generally received aspirin or no specific therapy.

The authors of this study warned that despite the excellent quality of information on treatment received and on outcome events, the study may not provide valid data on the (lack of) efficacy of warfarin anticoagulation in persons with a TIA. What do you believe is the primary basis for their cautious interpretation?

Answer 4.11 The authors likely are concerned with the presence of confounding—among patients with a TIA, warfarin anticoagulation treatment may have been selectively administered to those with an inherently greater stroke risk (perhaps based on age or the severity of the transient episode).

Question 4.12 Investigators at a clinic specializing in post-trauma care wished to determine whether persons who sustain acute neck trauma are at an increased risk of developing a diffuse pain syndrome (i.e., involving multiple parts of the body). They identified 102 patients in their clinic an average of 12 months after neck trauma, in whom 22 (21.6%) had symptoms that met the criteria for diffuse pain syndrome (DPS). In contrast, among 59 patients seen in the clinic an average of 12 months after a lower extremity fracture, only 1 (1.7%) reported symptoms consistent with DPS ($p = 0.001$). Based in large part on the strong association observed in this study, a review article concluded that the hypothesis that acute neck trauma can trigger DPS "meets established criteria for determining causality."

In this study, not all patients who sustained a neck injury were included, but rather just those who sought care in the clinic. Under what circumstance could this choice have led to a spurious exaggeration of the association between neck injury and DPS? (Assume that demographic and other risk factors for DPA are comparable between the patients with neck injury and those with lower extremity fracture.)

Answer 4.12 It may be that most or all patients with a lower extremity fracture require continuing care through 12 months postfracture, in other words, they will be seen even in the absence of any new condition. If the same were not true for neck trauma patients, then those seeking care from the clinic an average of 12 months later will be overrepresented by patients with new problems, including DPS. Only a cohort study that monitors the later DPS status of *all* victims of neck and lower extremity trauma can be trusted to provide a valid result.

Question 4.13 You are planning a cohort study of the possible influence of elective induction of labor at 39 weeks' gestation on the risk of several adverse maternal and fetal outcomes, including the need for cesarean section. As a basis for comparison to the women receiving elective induction, you are considering two possibilities:

(a) Women who delivered a child at 39 weeks who went into labor spontaneously and (like the women who were electively induced) did not have any indication for induction (e.g., preeclampsia, gestational diabetes)
(b) Women who at 39 weeks had not yet delivered and (as of that time) had no indication for induction.

Which of (a) or (b) would be the more likely to provide an unbiased estimate?

Answer 4.13 The correct answer is (b), because in this clinical scenario the alternative to a decision to electively induce labor at 39 weeks is to allow the pregnancy to continue. While some women whose delivery is not induced will indeed soon deliver on their own, others will remain pregnant at 40 weeks and beyond, during which time pregnancy complications often develop. These complications may lead to the decision to induce labor for medical reasons or to perform a cesarean section. Women experiencing these outcomes would be inappropriately excluded from comparison group (a). All of the women not undergoing elective induction of labor must be retained in the comparison population so as to obtain a valid assessment of the experience associated with this alternative course of action.

This has been noted by Caughey et al.,[27] who also provided some illustrative data on the occurrence of cesarean section:

*Induction of labor at 39 weeks?**	*% of women delivered by cesarean section*
Yes	14.3%
No	
Delivered at 39 weeks	9.1%
Delivered after 39 weeks	15.0%

*These generally were not elective inductions, but the presence of an indication for induction was adjusted for in the analysis.

The use of noninduced women who delivered at 39 weeks as a basis for the expected incidence of cesarean section would incorrectly suggest that induction of labor at 39 weeks predisposed to this outcome.

The didactic message: Comparison groups in cohort studies should not be subject to selection on the basis of events that occur after the time that the exposure has been sustained.

CHAPTER 5

Case-Control Studies

IN CASE-CONTROL STUDIES, persons with a given illness or injury (cases) are characterized as to whether or not they previously had sustained a given exposure (and degree of that exposure). Ideally, as a basis for comparison, we would like to determine the probability and degree of exposure in a sample of a population from which the cases were drawn. This ideal can be hard to achieve in practice, and it is common for the results of case-control studies to be ambiguous in their interpretation. Nonetheless, because case-control studies potentially are able to provide valid results, and because there are instances in which the only data relevant to a particular etiologic relation can be obtained from case-control studies, they are, warts and all, an essential item in the epidemiologist's toolbox.

This page intentionally left blank

Question 5.1 The following is excerpted from a letter to the editor of a medical journal:

> The fact that several relatives sometimes stutter has led others to assert that stuttering is inherited. Yet in forty years of experience I have met more stutterers who had no close relatives or ancestors who stuttered than those who had. It is my observation that most stutterers are hypersensitive persons, and I believe that hypersensitivity is acquired at an early age through the child's environment.

Does the information in the letter address the possible association between positive family history of stuttering and stuttering itself? If yes, how? If no, why not?

Answer 5.1 Even though only a minority of stutterers have a family history of this condition, that proportion may still be considerably in excess of the proportion present among nonstutterers. In the absence of data on the latter, no conclusions can be drawn regarding the presence or absence of an association.

Question 5.2 In a case-control study of pneumonia in infants in southern Brazil, the mothers of 152 cases and 2,391 controls were interviewed. A far higher proportion of cases than controls were not being breast-fed during the week prior to the date of onset of the case's illness (and the corresponding date for controls): odds ratio = 17, 95% confidence interval = 7.7–36.0.

A commentator on this study indicated that a "serious concern is the relatively small sample size, with the result that very few cases were exclusively breast fed." Do you agree that this is a "serious concern" when interpreting the results of the study? If yes, why? If no, why not?

Answer 5.2 No. The lower confidence limit of the odds ratio associated with not breast-feeding is 7.7; thus, there is little possibility that chance is solely responsible for the association. There would be concern over the sample size only if it were important to know more precisely where, within the range of 7.7–36.0, the true odds ratio lies.

Question 5.3 You are planning to conduct a case-control study that would examine a possible association between a genetic characteristic—called T1—and the incidence of lung cancer. About 10% of the population possesses T1. In this study, you have two laboratory methods that potentially could be used to measure a person's T1 status from a sample of his/her DNA. Neither is perfect. Method A will correctly categorize everyone who truly is positive for T1, but is expected to misclassify 5% of truly negative persons as being T1-positive. Conversely, method B will correctly classify everyone who is truly T1-negative, but will misclassify 5% of T1-positive persons as T1-negative. Which is the better choice of tests to minimize bias in the odds ratio relating T1 status and lung cancer?

a. Method A
b. Method B
c. Neither is a better choice than the other.

Explain your answer.

Answer 5.3 Method B will provide a less biased result, since its use will result in a relatively smaller proportion of misclassified individuals. As an example:

True T1 status	*Cases*	*Controls*	*Odds Ratio*
+	200	100	2.25
–	800	900	
Using Method A, 5% × 800 = 40 cases truly negative for T1 would be labeled as T1–positive, as would 5% × 900 = 45 controls.			
Method A: Observed T1 status			
+	200 + 40	100 + 45	1.64
–	800 – 40	900 – 45	
Using Method B, 5% × 200 = 10 cases truly negative for T1, would be labeled as T1-negative, as would 5% × 100 = 5 controls.			
Method B: Observed T1 status			
+	200 – 10	100 – 5	1.99
–	800 + 10	800 + 5	

Question 5.4 The following is excerpted from an article on potential risks associated with spray painting in the automobile industry[28]:

> In the case-control analyses, cases are defined as all lung cancer deaths among the automotive workers (n = 263). Controls are defined as those deaths due to circulatory disease or to accidents among those same workers, thus ensuring a valid representation of the population under investigation.

In this study, a history of spray painting was ascertained in cases and controls from company records.

a. What "population" is the one for which we would like to know the proportion of employees who worked as spray painters?
b. Apart from chance, how might the control group selected not accurately characterize the proportion exposed in that population?

Answer 5.4

a. The population is that of automotive workers at risk of death in the period during which case accrual took place.
b. If work as a spray painter is associated with the rate of death of the two causes selected—positively or negatively—then the proportion of spray painters in this control group will not be representative of that in the underlying population at risk.

Question 5.5 At the sexually transmitted infection (STI) clinic where you work, you are planning a case-control study of acute gonorrhea in men in relation to condom use. Because the cases of gonorrhea seen at the clinic are not derived from any defined population, you consider the use of male patients with an STI other than gonorrhea as a sampling frame for controls. This choice has the advantage of being feasible, and it is likely that the accuracy of information obtained on recent condom use would be similar between cases and controls. What do you believe to be the primary threat to the validity of this study if you were to select as controls men seen at the clinic for an STI other than gonorrhea?

Answer 5.5 The primary concern here is selection bias. Condom use (or the lack thereof) plausibly has a similar favorable impact on the incidence of both gonorrhea and other STIs. Therefore, the prevalence of condom use in STI controls may be considerably lower than in men in general. The use of such a control group could lead to a spuriously low estimate of the efficacy of condom use in the prevention of gonorrhea.

Question 5.6 This question is based on the following abstract.[29]

> *Study objective.* To investigate the association of alcohol use and night driving with traumatic snowmobile fatalities.
>
> *Design.* Case-control study.
>
> *Participants.* Traumatic deaths occurring while driving a snowmobile during the years 1985 to 1990 were reviewed. A sample of 1989 to 1990 fatal motor vehicle driver and motorcycle driver accidents were used as controls. Records were obtained from the provincial coroner.
>
> *Results.* One hundred eight snowmobile fatalities, 432 motor vehicle fatalities, and 108 motorcycle fatalities were included. Young men (mean age, 30 years) made up the snowmobile fatalities population, with weekend fatalities predominating (67%). Snowmobile fatalities were associated with use during times of suboptimal lighting (crude odds ratio, 1.9 [95% confidence interval, 1.1–3.3]; P <.01). Blood alcohol concentration exceeded provincial limits in 64% of cases. When snowmobile fatalities were adjusted for occurrence during suboptimal lighting conditions, only alcohol use was associated independently with fatal outcome (adjusted odds ratio, 4.3 [95% confidence interval, 2.5–7.0]; P <.0001).
>
> *Conclusion.* Drivers in snowmobile fatalities are associated with an approximately fourfold greater use of alcohol than are age- and sex-matched drivers in automobile and motorcycle fatalities.

Do you believe the control group chosen for this study led to bias in the estimate of the size of the association between fatal snowmobile trauma and alcohol use? If yes, why, and would an unbiased estimate be greater or smaller than that obtained by the authors? If not, why not?

Answer 5.6 The control group for this study is comprised of persons with alcohol-related causes of death, in other words, those due to motor vehicle and motorcycle accidents. Therefore, the observed odds ratio is almost certainly lower than the true one.

Question 5.7 Please answer the questions below after reading the following meeting abstract (modified).

> *Design*. Population based case-control study of drivers with known drug and alcohol concentrations who were involved in fatal crashes from October 2001 to September 2003. The cases were the 6766 drivers considered at fault in their crash; the controls were 3006 other drivers.
>
> *Results*. 681 drivers were positive for cannabis (cases 8.8%, controls 2.8%) (odds ratio 3.32, 95% confidence interval 2.63 to 4.18).
>
> *Conclusions*. Driving under the influence of cannabis increases the risk of involvement in a crash.

a. Blood specimens were readily available both for cases and controls, and obtaining such specimens on a representative sample of drivers would have posed formidable logistical difficulties. Nonetheless, this choice could have led to a result that was biased to at least some extent.
   i. Under what circumstance(s) would bias arise?
   ii. In what direction would the bias likely be operating, to falsely increase or falsely decrease the odds ratio? Why?

b. In this study, the population attributable risk % (PAR%, the percentage of this population's incidence of a fatal automobile crash attributable to the exposure in question) associated with positive detection of cannabis in blood was 6.1%. The corresponding PAR% for positive detection of alcohol in blood was 28.6%, despite the fact that the proportion of controls in whom each substance was detected was identical. What must be the explanation for the disparity between the size of the PAR% for cannabis and that for alcohol?

Answer 5.7

a. Bias would be present to the extent that the blood levels of tetrahydrocannabinol in drivers killed in auto crashes, who were judged not to be at fault in the crash, did not reflect that of drivers in general. This could occur if either one or both of the following were true:
   - The designation of "at fault" sometimes was in error, so that true cases were intermixed into the control group, giving that group a spuriously high proportion of apparent cannabis users.
   - Even if not "at fault," drivers who had consumed cannabis were impaired in their ability to avoid a crash.

In either circumstance above, the observed odds ratio associated with evidence of cannabis consumption would be spuriously low.

b. The odds ratio associated with alcohol consumption must have been considerably higher than that for cannabis consumption. (In fact, it was higher: 15.5 vs. 3.3.)

Question 5.8 This question pertains to the following abstract[30]:

> Acute influenza infection may be transiently associated with the risk of cardiovascular disease. We examined the association between influenza vaccination and incident myocardial infarction (MI) and stroke in a population-based case-control study. Case subjects were members of Group Health Cooperative (GHC) with incident MI or ischemic stroke during "flu season" (November-March) of 1992–1998. Control subjects were GHC members without history of MI or stroke who were frequency matched to case subjects by age, sex, and calendar year. The medical records of 584 case subjects with MI, 269 case subjects with ischemic stroke, and 1,415 controls were reviewed. Receipt of each year's influenza vaccine was not associated with risk of incident MI (odds ratio [OR] = 0.95, 95% confidence interval [CI]: 0.77, 1.17) or ischemic stroke (OR = 1.20, 95% CI: 0.91, 1.60) during the period of expected influenza activity. This study suggests that . . . influenza vaccination is not associated with a reduction in risk of first MI or ischemic stroke.

By restricting cases of MI and stroke to those that occurred during November-March, the investigators obtained a sample size that was considerably smaller than the one that would have included persons diagnosed with these illnesses in other months as well. What do you believe to have been the primary compensating advantage of the choice they made?

Answer 5.8 It was hypothesized that vaccination might prevent cases of MI and stroke that were precipitated by an influenza infection. Such infections occur primarily in the winter months. The ability of the study to find a true association would have been lessened if MI and stroke cases that were not precipitated by influenza, and thus did not have the potential to be prevented by vaccination, were included. That is, if there truly had been a beneficial impact of vaccination on the risk of MI and/or stroke, the observed odds ratio, using cases in all months, would have been closer to the null than that based on cases diagnosed in November-March.

Question 5.9 The following is excerpted from an article on mesothelioma in relation to employment[31]:

> In a case-control study, the occupational exposures of 259 mesothelioma patients were compared to those of an equal number of controls. Several occupations known to entail substantial exposure to asbestos were more common among cases than controls.

Longest held occupation	*Cases*	*Controls*
Insulator	47	13
Shipbuilder	31	21
Plumber	35	28
Furnace or boiler installer or repairman	21	10

> However, there were an identical number of cases and controls (15) who had engaged in brake lining work or repair, indicating that no increase in risk was associated with this type of employment.

Assume the following in this study:

a) Employment status was ascertained without error.
b) Cases and controls were completely comparable with regard to nonoccupational determinants of mesothelioma.
c) Among persons whose longest-held job was not one of the four listed in the above table, those who did brake lining work or repair were no more likely than other persons to also have engaged in one of those four occupations at some time in their lives.
d) Sampling variability is not an issue.

Do you agree with the author's conclusion above regarding the relation of mesothelioma occurrence to brake lining work or repair that was observed in their study? If yes, why? If no, why not?

Answer 5.9 No. To evaluate employment in brake lining work or repair as a possible risk factor for mesothelioma, it is necessary to estimate the mesothelioma incidence in persons with that exposure relative to the risk in a referent category comprised of persons not believed to be at increased risk of this disease. Thus, the analysis should be restricted to persons who had not been employed in furnace and boiler work, insulation, and so forth. If persons who had been insulators, and so forth, had been excluded from the analysis, the referent category would be reduced by 47 + 31 + 35 + 21 = 134 cases and by 13 + 21 + 28 + 10 = 72 controls. The odds ratio associated with employment in brake lining and repair would be as follows:

Brake work	*Cases*	*Controls*
+	15	15
–	259 – 15 – 134 = 110	259 – 15 – 72 = 172

$$\text{Odds ratio} = \frac{15}{110} \div \frac{15}{172} = 1.56$$

The authors' interpretation of the results of their study is incorrect.

(It turns out that assumption (c) above is almost certainly not a valid one, i.e., a relatively high proportion of men whose longest held job involved brake lining work or repair indeed had been employed at other times in a high risk occupation. When in a reanalysis attention was restricted to men who had no such history,[34] no greater proportion of mesothelioma cases (1/33) than controls (9/171) had been employed to do brake work.)

Question 5.10 Epidemiologists in California observed that about 65% of infants who died of sudden infant death syndrome (SIDS) typically were put to sleep in the prone position, in contrast to 60% of control infants. Most other studies of SIDS and sleeping position have found a considerably greater case-control difference. Commenting on this result, an editorialist wrote, "One reason that prone sleeping may not have been observed as a strong risk factor for SIDS [in this study] is that it is difficult to measure risk for a characteristic present in 60 percent of the population." Do you agree with this assertion? If yes, why? If not, why not? If not, why not?

Answer 5.10 No, you do not agree.

In terms of the frequency of the exposure, the only circumstance in which case-control studies find "it is difficult to measure risk" is when the exposure either is extremely common (e.g., > 95%) or extremely uncommon (e.g., < 5%). (In those instances, study power is reduced for a given sample size.)

An exposure frequency of 60% is nowhere near these values.

Question 5.11 You have been sent a manuscript to review. It describes the results of a case-control study of gastric ulcer in relation to prior use of nonsteroidal anti-inflammatory drugs (NSAIDS). Cases in this study were all 225 patients diagnosed with a first gastric ulcer between 1982 and 1985 in a given population. All diagnoses were made by endoscopy that had been performed at specialized centers for this purpose. It is believed that gastric endoscopy was not done elsewhere in that part of the world. Community controls were selected randomly from electoral rolls for 1982–1985. After conducting a screening interview with potential control subjects (matched to cases on the basis of sex, age, and area of residence), those with a history of gastric ulcer were excluded. In addition, about 13% of potential controls with dyspepsia were excluded as well: The authors were concerned that: (a) some persons with this symptom might have a gastric ulcer that had not yet been diagnosed; and (b) because NSAID use is a likely cause of dyspepsia, the occurrence of dyspepsia would be considerably more common in NSAID users than in other persons.

Assume that only a very small percentage of those potential controls with dyspepsia actually had a gastric ulcer. Do you agree with the authors' choice to exclude persons with dyspepsia? If yes, why? If no, why not?

Answer 5.11 You disagree. The controls should be a sample of the at-risk population. Cases were included whether or not they had dyspepsia, so persons with dyspepsia were part of the at-risk population. By restricting controls to persons free of dyspepsia, the proportion of persons with a history of use of NSAID would be spuriously low. This would lead to an overestimate of the odds ratio relating a history of NSAID use to the occurrence of gastric ulcer.

Question 5.12 The following is an excerpt from an abstract of a case-control study[32]:

> Experimental and epidemiologic evidence has suggested that phenacetin use increases the risk of transitional cell cancers of the urinary tract. The drug is no longer marketed but a commonly used metabolite, acetaminophen, has been linked recently to an increased risk of renal cancer. We assessed the relation of acetaminophen use to the risk of transitional cell cancer of the urinary tract with data from a hospital-based study of cancers and medication use conducted from 1976–96 in the eastern United States. We compared interviews with 498 cases of transitional cell cancer with those of 8,149 noncancer controls, and controlled confounding factors with logistic regression. For transitional cell cancer, the relative risk (RR) estimate for regular acetaminophen use that had begun at least a year before admission was 1.1 (95% confidence interval (CI = 0.6–1.9). RR estimates for use that lasted at least five years, and for non-regular use, were also close to 1.0. Our results suggest that acetaminophen, as used in present study population, does not influence the risk of transitional cell cancer of the urinary tract.

The authors stated that, in choosing noncancer controls they included only persons hospitalized for conditions that "were judged to be unrelated to acetaminophen use. For example, we did not include patients admitted for gastric or duodenal ulcers, because such persons might have used acetaminophen preferentially to aspirin for pain relief." Do you agree with this strategy? Why? If not, why not?

Answer 5.12 Yes, the strategy is reasonable. By not including patients who were hospitalized because of a condition that is an indication for acetaminophen use, the authors assembled a group of hospitalized patients whose exposure history possibly could reflect that of the source population of the cases. However, it is plausible that a relatively high proportion of hospitalized persons have a history of analgesic use, whether aspirin or acetaminophen. If this is true, the proportion of controls selected for this study who previously had taken acetaminophen would exceed that of the at-risk population, leading to a falsely low odds ratio.

Question 5.13 The following question is based on an excerpt of the abstract of an article "Patent foramen ovale and cryptogenic stroke in older patients"[33]:

> We prospectively examined 503 consecutive patients who had had a stroke, and we compared the 227 patients with cryptogenic stroke and the 276 control patients with stroke of known cause. We examined the prevalence of patent foramen ovale in all patients, using transesophageal echocardiography.
>
> The prevalence of patent foramen ovale [a congenital heart defect] was significantly greater among patients with cryptogenic stroke than among those with stroke of known cause, for both younger patients (43.9% vs. 14.3%; odds ratio, 4.70; 95% confidence interval [CI], 1.89 to 11.68; P<0.001) and older patients (28.3% vs. 11.9%; odds ratio, 2.92; 95% CI, 1.70 to 5.01; P <0.001).

In theory, the control group against which cases of cryptogenic stroke ought to be compared for the prevalence of patent foramen ovale is a sample of persons who are demographically similar to the cases but otherwise unselected. Do you believe that the control group actually chosen in this study correctly produced a positive association between patent foramen ovale and the occurrence of cryptogenic stroke? If yes, why? If no, why not?

Answer 5.13 The control group used—patients with a stroke of "known" cause—will provide a valid result to the extent that the prevalence of patent foramen ovale (PFO) in these persons reflects that of the underlying population from which the cases of cryptogenic stroke arose. *If* the attribution of a known cause had been correct, then a valid result likely would have been obtained, given that there is no reason to believe that the presence of PFO influenced the development of a stroke in these persons. To the extent that some of the cases with a "known" cause were in truth cryptogenic ones, then the observed association between PFO and cryptogenic stroke actually would be spuriously small. (However, the prevalence of PFO in the controls chosen—11.9%–14.3%—was not elevated relative to that expected based on autopsy studies.)

CHAPTER 6

Multiple Causal Pathways and Effect Modification

AN IMPORTANT THREAT to an epidemiologic study's sensitivity in identifying a genuine exposure–disease association is error in the measurement of the exposure and/or the illness outcome. Confounding can be another threat. A third, not as widely appreciated as measurement error and confounding, is a study's failure to take into account the presence of causal pathways leading to disease other than the one under consideration. The means by which these other pathways can be accommodated vary from study to study. Sometimes, illness outcomes can be subdivided based on the presence or absence of a manifestation of the condition in question, for example, estrogen receptor-positive breast cancer from estrogen receptor-negative breast cancer in studies of hormonal exposures. Alternatively, persons in whom a known potent etiologic agent is present can be omitted from consideration, so as to allow the influence of an exposure acting to cause disease through a different means to be seen (e.g., in studies of mental retardation, when children with microcephaly are excluded when studying the potential influence of postnatal exposure to low levels of metals in the environment). However, interpreting the results of analyses based on subgroups of the study population is not always straightforward—as we'll see in some of the following examples.

This page intentionally left blank

Question 6.1 You come across an abstract of an article in a medical journal, which in part reads as follows:

> In our study of women diagnosed with endometrial cancer, possible risk factors were identified through personal interviews. Among women with endometrial cancer who had and had not taken unopposed estrogen therapy prior to the time the tumor was diagnosed, there were no differences with regard to parity [i.e., the number of children they had borne]. The results of this study do not support the hypothesis that parity is differentially associated with endometrial cancer that is and is not related to exogenous estrogens.

Based on the above summary, what do you believe to be the main limitation of this study in evaluating whether parity has a different relationship to the etiology of endometrial cancer that is and is not related to unopposed estrogen use?

Answer 6.1 The differential relationship between a particular characteristic of a woman's reproductive history and endometrial cancer, according to a history of estrogen use, can be discerned only by measuring the size of the association between that characteristic and endometrial cancer in estrogen users and then again in nonusers. Since no comparison group is mentioned, the presence of an association cannot be assessed in either group of women.

Question 6.2 Persons with a factor V Leiden mutation are resistant to the anticoagulant effect of activated protein C. The following table describes the incidence of first venous thrombosis (VT) in women ages 15 to 49 years, according to presence of the factor V Leiden mutation and the use of oral contraceptives (OC):

	Cases of VT	*Person-years*	*Incidence of VT per 10,000 person-years*
Factor V Leiden negative			
No OC use	36	437, 870	0.8
Current OC use	84	275, 858	3.0
Factor V Leiden positive			
No OC use	10	17, 515	5.7
Current OC use	25	8757	28.5

Assume that the incidence of first and recurrent VT bear a similar relation to oral contraceptive use and the factor V Leiden mutation. In a 15- to 49-year-old woman who develops VT, do the data presented above argue that her factor V Leiden status should be considered in counseling about her future method of contraception?

Answer 6.2

Factor V Leiden	*Incidence of VT per 10,000 person-years*	*Rate ratio*	*Rate difference per 10,000 person-years*
Negative			
No OC	0.8		
Current OC	3.0	3.7	2.2
Positive			
No OC	5.4		
Current	28.2	5.0	22.8

Yes. Although OC use is associated with an increased incidence of VT regardless of factor V Leiden mutation status, the recommendation is supported by the considerably larger absolute increase in the rate of VT associated with OC use in factor V Leiden-positive than in factor V Leiden-negative women: 22.8 per 10,000 person-years versus 2.2 per 10,000 person-years.

Question 6.3 A study observed that, for a given history of cigarette smoking, the relative risk for lung cancer among women was 1.2- to 1.7-fold greater than it was in men. The disparity between the sexes was well beyond that expected by chance, and led the authors to conclude that women were more susceptible than men to respiratory carcinogenesis from cigarette smoke.

A subsequent letter to the editor of the journal in which the article was published suggested that in order to use the data from the study to infer differential susceptibility between the sexes, it would be necessary to know the incidence of lung cancer in men who had not smoked and in women who had not smoked. Why would this additional information be useful?

Answer 6.3 If the rate in female nonsmokers were lower than in male nonsmokers, the same added risk associated with cigarette smoking in the two sexes would produce a larger relative risk in women.

For example:

	*Incidence in nonsmokers**	*Incidence in smokers**	*Attributable risk**	*Relative risk*
Men	10	30	20	3
Women	5	25	20	5

*Rate per 100,000 person-years.

Therefore, depending on how *differential susceptibility* is defined—greater relative risk versus greater attributable risk—women and men may or may not differ in terms of the impact of smoking on their incidence of lung cancer.

Question 6.4 This question pertains to the following excerpt of an abstract[35]:

> *Background*. Elderly people who have a fracture are at high risk of another. Vitamin D and calcium supplements are often recommended for fracture prevention. We aimed to assess whether vitamin D3 and calcium, either alone or in combination, were effective in prevention of secondary fractures.
>
> *Methods*. In a factorial-design trial, 5,292 people aged 70 years or older (4481 [85%] of whom were women) who were mobile before developing a low-trauma fracture were randomly assigned 800 IU daily oral vitamin D3, 1000 mg calcium, oral vitamin D3 (800 IU per day) combined with calcium (1000 mg per day), or placebo. Participants who were recruited in 21 UK hospitals were followed up for between 24 months and 62 months. Analysis was by intention-to-treat and the primary outcome was a low-energy fracture.

In their trial, the investigators identified 698 participants in whom a "low-energy" fracture occurred. Not included in the analysis were an additional 34 fractures that resulted from substantial trauma, for example, a fracture sustained in an automobile crash.

The disadvantage of not including the 34 fractures involving more than a low level of trauma is a reduced sample size (by about 5%). What do you believe to be the main advantage of this choice?

Answer 6.4 The investigators believed that the relative influence of one or more of the intervention measures on the risk of fracture could differ depending on the level of trauma to which a participant was exposed. If very little trauma were present, the relative benefit might be great; in the presence of substantial trauma, there might be but little benefit from the intervention. Therefore, to maximize the sensitivity of the study to observe any relative change in fracture occurrence associated with one or more treatments, the analysis excluded persons with fracture in whom another factor, substantial trauma, likely played a causal role.

Question 6.5 You observe that among persons 20 years of age or older who developed a particular infectious disease, 32% had been vaccinated against that disease, in contrast to 16% of persons under 20 years with the disease. Is this necessarily evidence of greater vaccine efficacy in younger persons?

Answer 6.5 No, such a conclusion would not be justified. In order to assess vaccine efficacy in either age group, it would be necessary to estimate the percentage of the population at risk who had been vaccinated.

	<20		*≥20*	
Vaccinated	*Cases*	*Population*	*Cases*	*Population*
Yes	32%	?	16%	?
No	68%	?	84%	?

Question 6.6 In a study of the possible adverse effect of the drug rosuvastatin on the incidence of venous thromboembolism (VTE),[36] the investigators separately analyzed "provoked" and "unprovoked" cases. "Provoked" cases were defined as those in which a strong risk factor had been present, such as recent surgery, immobility or metastatic cancer. What do you believe to have been the rationale for making this separation?

Answer 6.6 If rosuvastatin acted together with other VTE risk factors in a causal pathway, the size of the association might have been expected to be particularly great for "provoked" cases. Alternatively, if the means by which rosuvastatin predisposed to VTE were independent of other risk factors, the relative risk associated with its use would be expected to be greatest for unprovoked VTE. An analysis that examines each type of VTE separately enables the evaluation of either possibility.

Question 6.7 The data presented in the table below appeared in a publication of the results of a case-control study on the risk of endometrial cancer in relation to physical activity and obesity.[37] Based on the data in the table, the authors stated that "the increase in risk associated with obesity [BMI ≥30] was much lower in active women (OR = 1.57) than in women with low physical activity (OR = 3.10)." What is a more accurate way of quantifying the difference in the relative risk of endometrial cancer associated with BMI ≥30 within the two categories of physical activity? Why might the approach used by the authors be misleading?

Lifetime physical activity and endometrial cancer risk, by BMI

	Average lifetime physical activity					
	Low			*High*		
BMI	*Cases*	*Controls*	*OR (95% CI)*†	*Cases*	*Controls*	*OR (95% CI)*†
<25	105	100	1.0	59	74	0.95 (0.57–1.58)
25–29	81	72	1.0	52	84	0.58 (0.34–0.99)
≥30	113	49	1.0	62	64	0.50 (0.28–0.91)
<25	105	100	1.0	59	74	0.90 (0.56–1.43)
25–29	81	72	1.39 (0.88–2.20)	52	84	0.85 (0.52–1.38)
≥30	113	49	3.10 (1.91–5.01)	62	64	1.57 (0.94–2.62)

*Below and above the median.

†Odds ratio (OR) and 95% CI were adjusted for age, race/ethnicity, education, family history of endometrial cancer, age at menarche, full-term pregnancies, duration of oral contraceptive use, duration of hormone therapy use, menopausal status, and height.

Answer 6.7 We wish to know whether the size of the OR associated with a high BMI on risk differs according to physical activity. Among less active women, the OR associated with obesity is 3.1. Among more active women, it is 1.57/0.90 = 1.74. (The OR of 1.57 incorporates *both* the elevated risk associated with obesity and the reduced risk associated with activity.)

Question 6.8 In a cohort study, postmenopausal women who were taking combined estrogen-progestogen hormone therapy at baseline had a 5-year cumulative incidence of breast cancer of 5.6/1,000 if their mother had a history of breast cancer, and 2.2/1,000 if not. Among hormone nonusers, the corresponding cumulative incidences were 5.1/1,000 and 1.7/1,000, respectively.

a. Among women with a maternal history of breast cancer, what was the relative risk of breast cancer associated with use of hormone therapy at baseline? The risk difference?
b. What were these same measures of excess risk in women without a maternal history of breast cancer?
c. The two means of assessing the potentially modifying influence of maternal breast cancer on the association between hormone use and risk of breast cancer do not produce the same qualitative result. Why is this?

You are trying to decide whether to differentially counsel postmenopausal women with and without a maternal history of breast cancer with regard to the impact of hormone therapy (perhaps in terms of the frequency of breast screening exams). Which contrast do you believe to be the most relevant for this purpose, the one between the two relative risks or that between the two risk differences? Explain.

Answer 6.8 a.

Maternal history	*Relative risk*	*Risk difference (per 1,000)*
+	5.6 per 1000/5.1 per 1000 = 1.1	5.6 – 5.1 = 0.5
–	2.2 per 1000/1.7 per 1000 = 1.3	2.2 – 1.7 = 0.5

b. In this instance, the same absolute increase in risk associated with hormone therapy (0.5/1,000) is acting upon a smaller background risk in women without a positive maternal history, producing a greater relative change in their risk (1.3) than the relative change (1.1) in women with a positive maternal history.

c. The size of the absolute change in risk associated with hormone therapy is what bears on personal decision making. Because the additional risk associated with hormone therapy is 0.5/1,000 irrespective of maternal history, counseling should not be differential between the two groups of women with regard to actions that might be taken.

Question 6.9 A study was conducted in sub-Saharan Africa among heterosexual couples in whom just one member was infected with HIV.[38] During the course of a 24-month follow-up, about 10% of participants who were infected at baseline initiated antiretroviral therapy. The study observed that among persons who were HIV-negative at baseline, the acquisition of an HIV infection that was phylogenetically linked to that of their HIV-infected partner was relatively less common among those whose partners had received antiretroviral therapy (0.37 vs. 2.24 per 100 person-years).

Had the study endpoint been *any* new HIV infection, irrespective of its linkage to the HIV strain of the initially infected partner (so as to include HIV infections that arose from sexual contact with other persons), what would have been the expected impact (if any) on the following measures of association:

a) Rate difference
b) Rate ratio

Explain.

Answer 6.9 If receipt of antiretroviral therapy by one's HIV-infected partner were unrelated to the likelihood of acquiring an HIV infection from a different partner, then the size of the absolute increase in the incidence of any new HIV infection above that of the "linked" infections would be the same whether or not antiretroviral therapy had been administered. In this circumstance, the rate difference would be unchanged, but the rate ratio associated with the partner's receipt of antiretroviral therapy would be closer to the null.

For example, assume an incidence of "unlinked" HIV infection in initially HIV-negative persons of 0.20 per 100 person-years:

Type of HIV infection	*Antiretroviral treatment*		*Rate difference (per 100 person-years)*	*Rate ratio*
	Yes	*No*		
Linked	0.37	2.24	1.87	6.1
Unlinked	0.20	0.20	0	1.0
Total	0.57	2.44	1.87	4.3

The presence of a second causal factor leading to disease (in this instance, sexual contact with another HIV-infected person)—one that does not join with the first factor in an etiologic pathway and is not accounted for in the study design and/or analysis—will lead to an attenuation of the estimated rate ratio of the disease associated with the first causal factor.

Question 6.10 The following letter, somewhat paraphrased, appeared in a medical journal:

> Bailar and Gornik report that the age-adjusted rate of mortality from all cancers in the United States declined by 1 percent from 1991 through 1994. Our estimate for the same interval is 2.2 percent. The discrepancy in the two figures stems from the use of different populations for age adjustment. Bailar and Gornik used the relatively elderly 1990 U.S. population and by doing so, minimized striking reductions in mortality that occurred among young and middle-aged persons. We used the U.S. (relatively younger) 1940 population, which reveals the full downturn in cancer-related mortality.

Suppose that you also are interested in quantifying the change in U.S. cancer mortality during 1991-1994, and have access to age-specific U.S. mortality rates for those years. In your analysis, would there be any virtue in presenting adjusted rates within specific age categories (e.g., "young and middle-aged" versus "older")? If yes, why? If no, why not?

Answer 6.10 Yes, there would be value in presenting age-specific rates. During the 4-year period under study there was, apparently, a "striking reduction" in cancer mortality in the United States in young and middle-aged persons, but not in older persons. Only the presentation of age-specific differences over time can illustrate this difference in trends. A measure such as a difference in adjusted rates across all ages will provide a summary that may not apply to any individual age group. Furthermore, as is clear from the letter to the editor, the size of the mortality trend will be influenced by the arbitrary choice of the age distribution to be used as the standard.

Question 6.11 The following is an abstract of an article "Occupational asbestos exposure and the incidence of non-Hodgkin lymphoma of the gastrointestinal tract: An ecologic study."[39]

> *Purpose*. A previous case-control study observed a strong association between occupational exposure to asbestos and the incidence of non-Hodgkin lymphoma of the gastrointestinal tract (GINHL). To test this hypothesis we sought to determine whether the geographic pattern of the incidence of GINHL in the US has paralleled that of mesothelioma.
>
> *Methods*. Using data obtained from the nine US regions participating in the National Cancer Institute's Surveillance, Epidemiology, and End Results program, we examined the incidence of malignancies among men ages 50 to 84 years between 1973 and 1984.
>
> *Results*. The rates of mesothelioma, but not of GINHL, were about two times higher in the areas of Seattle and San Francisco than in the other regions. Overall, there was no correlation between the rates of mesothelioma and of GIHNL (Person correlation coefficient—0.12, $\rho = 0.77$).
>
> *Conclusions*. This ecologic study finds no support for the hypothesis that occupational asbestos exposure is related to the subsequent incidence of GINHL.

In their analysis, the authors paid particular attention to rates in men (in whom the likelihood of prior occupational exposure was far greater than in women) and to rates in 50- to 84-year-olds (to allow for a potentially long induction period). And, even though the Surveillance, Epidemiology, and End Results program had data available through the 1990s, the authors confined their analysis to cancer incidence through just 1984. What do you believe to have been their reason for this latter choice? (Hint: The presence of HIV infection strongly predisposes to the development of GINHL.)

Answer 6.11 GINHL has a number of causal pathways that may lead to its occurrence. One of these involves HIV infection. If occupational asbestos exposure adds to a person's risk of GINHL to the same extent whether HIV infection is present or not, then the relative increase in risk associated with asbestos exposure will be greatest in HIV-uninfected persons. Therefore, the most sensitive assessment of the potential role of asbestos is to exclude cases related to HIV. One way of accomplishing this is to restrict the time period being considered to that before HIV infection was widespread, in other words, prior to 1985.

(Because the prevalence of HIV infection varies geographically across the United States, failure to control for HIV infection in this way also could lead to confounding. For example, GINHL rates in San Francisco during the last decade of the twentieth century might be high relative to other parts of the United States because of the relatively high prevalence of HIV infection there, and not because of a higher degree of occupational asbestos exposure.)

Question 6.12 A study sought to determine if pregnancy intendedness is associated with intimate-partner physical violence, and to identify factors that modify this association. Three to 6 months after delivery, the investigators mailed a questionnaire to a population-based sample of 12,612 mothers of infants born in four states.

Some of the results of this study are shown in the following table:

Percentage of women experiencing physical violence during pregnancy, by pregnancy intendedness and education

	Unwanted		*Intended*	
Education (y)	*%*	*(95% CI)*	*%*	*(95% CI)*
All	12.1	(8.8–15.6)	3.2	(2.4–4.0)
<12	18.6	(9.6–27.6)	7.0	(4.1–9.9)
12	10.6	(5.9–15.3)	3.7	(2.5–4.9)
>12	9.2	(2.7–15.7)	1.3	(0.7–1.9)

The authors stated that the risk for experiencing violence in women who had an unwanted pregnancy, relative to the risk in women with an intended pregnancy, was particularly high among women with more social advantage. For example, the relative prevalence among women with fewer than 12 years of education was 2.6, whereas the corresponding relative prevalence in women with >12 years of education 7.1.

In discussing the findings, the authors put forth some possible explanations "for the interaction between pregnancy intendedness, social status, and physical violence." Could it be argued that, apart from the issue of chance (i.e., sampling variability), there is *no* interaction to account for? If yes, why? If no, why not?

Answer 6.12 The investigators used a relative measure of association to assess the possibility of an "interaction" between education and pregnancy intendedness as a predictor of physical violence. However, the data suggest that, in absolute terms, the increased risk of experiencing violence associated with an unwanted versus an intended pregnancy is about the same for all levels of education. If anything, the observed difference in prevalence was greater in women with <12 years of education (18.6% – 7.0% = 11.6%) than in women with >12 years of education (9.2% – 1.3% = 7.9%).

Question 6.13 The data presented in the following table are from a hypothetical study of perinatal death among twins in relation to order of delivery.*†

Gestational age (weeks)	*No. of twin pregnancies*	*No. of perinatal deaths in:*		
		first-born twin (a)	*second-born twin (b)*	*b/a*
24–27	703	359	381	1.06
28–31	1,371	98	141	1.44
32–35	2,897	74	164	2.22
≥36	2,935	33	124	3.76

Would it be reasonable to infer from these results that the increase in risk of perinatal death in the second-born twin is a particular concern in a term or a near-term pregnancy, and less so in a pregnancy that ends well prior to term? Explain.

* Infants born to women undergoing a planned caesarean section are excluded.
† Restricted to pairs in which at least one twin survived.

Answer 6.13 At gestational ages 24–27 weeks, the risk of perinatal death in first-born twins is 359/703 = 51/100, and that in second-born twins is 381/703 = 54/100. The difference in risk of death between second- and first-delivered twins at other gestational ages also is about 3 per 100. Therefore, although the relative difference in risk of death differs across gestational age categories, the added risk among second-delivered twins is of similar concern irrespective of gestational age.

	Risk of perinatal death (per 100)			
Gestational age (weeks)	*First-born twin*	*Second-born twin*	*Relative risk*	*Risk difference (per 100)*
24–27	51.1	54.2	1.06	3.1
28–31	7.1	10.3	1.45	3.2
32–35	2.6	5.7	2.19	3.1
≥36	1.1	4.2	3.82	3.1

Question 6.14 The following is adapted from a report of a case-control study of primary liver cancer in relation to serum levels of retinol, in which serum samples had been obtained prior to the diagnosis of cancer.[40]

> Men with low prediagnostic serum retinol levels had a relatively high risk of liver cancer. A statistically significant interaction was observed between retinol levels and hepatitis B surface antigen (HBsAg) seropositivity on cancer risk: HBsAg-positive men in the lowest third of the distribution of serum retinol had greater than a 70-fold higher risk than HBsAg-negative men in the highest third of the distribution of serum retinol (p for interaction = .018).

From the information provided, can you determine the nature of the interaction between retinol levels and HBsAg status with regard to the incidence of liver cancer? If yes, what is it? If not, what additional information from the study would you need?

Answer 6.14 No, the nature of the interaction cannot be determined. In each of the following examples, assume (for simplicity) that the odds ratio associated with the presence of both low retinol levels and HBsAg+ relative to high retinol levels and HBsAg– is not 70, but instead is 100 (200/100 ÷ 20/1000), based on the hypothetical data shown below.

The risk of liver cancer associated with low retinol levels might be particularly great in men who are HBsAg positive:

	HBsAg+			*HBsAg–*	
	Case	*Control*		*Case*	*Control*
Low retinol	200	100	Low retinol	700	1000
High retinol	10	100	High retinol	20	1000
Odds ratio		20			3.5

Or, it might not:

	HBsAg+			*HBsAg–*	
	Case	*Control*		*Case*	*Control*
Low retinol	200	100	Low retinol	700	1000
High retinol	10	100	High retinol	20	1000
Odds ratio		20			35

Only in the first example is the odds ratio associated with low retinol levels greater in HBsAg+ men than HBsAg– men. Therefore, what's needed to understand the nature of the interaction is the size of the odds ratio (or the risk difference, if examining a deviation from additivity) associated with retinol levels within categories of HBsAg seropositivity.

Question 6.15 The following data come from a case-control study of upper gastrointestinal bleeding (UGIB) in relation to prior use of nonsteroidal anti-inflammatory drugs (NSAIDS).

History of ulcer	*NSAID use*	*Cases*	*Controls*
No	No	607	15,242
No	Yes	171	597
Yes	No	405	1,430
Yes	Yes	106	164

Do the data from this study suggest that, when deciding to initiate treatment with an NSAID, the risk of UGIB should weigh less heavily as a potential adverse effect in persons with a history of an ulcer than in other persons? Explain your answer.

Answer 6.15 The decision to use an NSAID (or any other drug) should be based in part on a comparison of the size of the added benefits and added risks that such use would entail. From the data obtained in this case-control study, it is possible to estimate the relative risk of UGIB (by means of the odds ratio) but not the added ("attributable") risk. Nonetheless, the relative size of the added risk associated with NSAID use in persons with, and in persons without, a history of ulcer can be calculated, as shown below:

Ulcer	*NSAIDS*	*Cases*	*Controls*	*Odds ratio*[a]	*Odds ratio*[b]	*Relative added risk*
No	No	607	15,242	1	1	7.19 – 1 = 6.19
No	Yes	171	597	7.19	7.19	
Yes	No	405	1,430	1	7.11	16.23 – 7.11 = 9.12
Yes	Yes	106	164	2.28	16.23	

[a] Separate referent category of nonusers of NSAIDs for persons with and without a history of ulcer.
[b] Common referent category = Persons with neither a history of ulcer nor of NSAID use.

The relative added risks associated with use of an NSAID obtained above have no units, but nonetheless can be meaningfully compared between persons with and without ulcers. Since the added risk of UGIB incurred as a result of NSAID use is similar in the two groups (if anything, it is greater in those *with* ulcers [i.e., 9.12 vs. 6.19]), it should receive equal weight by potential NSAID users irrespective of ulcer history. The fact that the first odds ratio is so much smaller in patients with ulcers than in patients without ulcers is attributable to the higher underlying rate of UGIB in ulcer patients.

CHAPTER **7**

Screening

AT A POPULATION LEVEL, the degree to which a screening test can lead to improved health outcomes is related to:

a. The proportion of the population with the condition (or predictor of that condition) that the test seeks to detect;
b. The sensitivity of the test in identifying the condition or predictor;
c. A low frequency of false positive tests, or a low frequency of adverse health outcomes associated with the consequences of a false positive test; and
d. The efficacy of treatment of persons who screen as positive.

Often, elements a-d above are addressed in separate studies. Occasionally, though, especially when we don't have available to us the experience of untreated persons who have tested positive, a single study seeks to examine the aggregate impact of two or more elements (e.g., a comparison of cancer mortality in screened and unscreened persons). The exercises in this chapter consider studies of the individual elements as well as those that deal with several of these at once.

This page intentionally left blank

Question 7.1 In 132 patients with cirrhosis of the liver, serum levels of alpha-L-fucosidase were measured and the patients were followed for 8 years for the occurrence of liver cancer. In 12 of them, levels of this enzyme were "high." Liver cancer was diagnosed in 19 patients, and in 3 of the 19, serum alpha-L-fucosidase levels were "high."

a. From the data provided, calculate the
   1. Sensitivity
   2. Specificity
   3. Predictive value of a positive test
   4. Predictive value of a negative test

   of a "high" level of serum alpha-L-fucosidase in patents with cirrhosis for predicting liver cancer during the ensuing eight years.
b. There was a "significant increase" 6 to 9 months before evidence of liver cancer in 7 of the 16 patients who had low serum levels of alpha-L-fucosidase at enrollment. What other information is needed before concluding that a change of this sort is useful in predicting the presence of liver cancer?
c. The authors of this paper concluded, "We recommend the measurement of this enzyme activity in surveillance programs for cirrhotic patients." Even if the serum alpha-L-fucosidase level in a person with cirrhosis perfectly discriminated between those who did and did not develop liver cancer, is it possible that this recommendation could be misguided? How?

Answer 7.1

"High activity"	*Cancer*	*No cancer*	*Total*
Yes	3	9	12
No	16	104	120
	19	113	132

a. 1. Sensitivity = 3/19 = 0.16
   2. Specificity = 104/113 = 0.92
   3. PV+ = 3/12 = 0.25
   4. PV– = 104/120 = 0.87

b. Other information needed: In what proportion of patients who did not develop liver cancer was there a "significant increase" in levels?

c. The recommendation could be misguided if, on average, early recognition of a liver tumor, by means of this test, did not result in an improved outcome.

Question 7.2 The following appeared in the "News" section of the *Journal of the National Cancer Institute*, May 17, 2000:

> **Some Promising Biomarkers for Cancer**
>
> - **LPA (lysophosphatidic acid).** "LPA is probably the most accurate marker we have for detection of early stage ovarian cancer," said Northwestern University's David Fishman, M.D., who is heading a multi-center study of the marker. A 1998 report from the Cleveland Clinic found 9 of 10 women with stage 1 disease, 24 of 24 with advanced disease, and 14 of 14 with recurrent ovarian cancer had elevated blood LPA levels. In contrast, just 5 of 48 controls had elevated LPA. A growth factor, LPA is not generally present in normal ovary cells.

Based on the above information, you believe it *un*likely that blood LPA levels will be of practical use in the early detection of ovarian cancer. What is your reasoning?

Answer 7.2 If the prevalence of ovarian cancer among screened women is low, the number of false positive tests—many of which would lead to a surgical procedure to document the absence of ovarian cancer—would greatly exceed the number of true positives. If the prevalence of cancer is 1/2,000, for example, these would be the expected results in 20,000 screened women:

	Ovarian cancer		
LPA	*Yes*	*No*	*Total*
Positive	10	5/48 (19,990) = 2,082	2,092
Negative	0	43/48 (19,990) = 17,908	19,990
	10	19,990	20,000

The PV+ would be 10/2,092 = 0.005, very likely too low to warrant use of LPA for early detection. For conditions whose prevalence is relatively low, this question illustrates the strong influence of a test's specificity on the predictive value of a positive test result.

Question 7.3 A test was developed (based on levels of a certain peptide in blood) to identify persons with congestive heart failure. After studies revealed that a negative test predicts the absence of heart failure virtually 100% of the time, the manufacturer of the test concluded that a positive test is "an unambiguous warning sign" for the presence of congestive heart failure. You disagree. Why?

Answer 7.3

	Congestive heart failure		
Test Results	*Yes*	*No*	*Total*
+	A	b	a + b
–	c (0%)	d (100%)	c + d

In order for a positive test to represent "an unambiguous warning sign," its positive predictive value (a/(a + b)) would have to be high. However, the presence of a negative predictive value (d/(c + d)) of 100% does not speak to this issue.

Question 7.4 The questions below pertain to the following abstract (abridged):

> **Survival of Women Ages 40-49 Years with Breast Carcinoma according to Method of Detection**
>
> *Methods*. Women ages 40-49 years diagnosed with invasive breast carcinoma between 1986 and 1992 were identified. Measures of tumor size, lymph node status, and overall survival were compared with breast carcinoma patients whose tumors were detected by breast self-exam (BSE), clinical breast exam (CBE), patient incidental finding (PI), or mammography.
>
> *Results*. Mean tumor size among women in the mammography group was smaller than that among women in the BSE, CBE, and PI groups ($P < 0.002$).Tumors detected by mammography were significantly more likely to be localized than those detected by other methods ($P < 0.0001$). Patients whose tumors were detected by mammography had significantly better survival than patients in the other detection method ($P < 0.0001$).
>
> *Conclusions*. Women ages 40-49 years whose invasive breast carcinoma is detected by mammography have significantly smaller tumors, more localized disease, and may have a lower risk of mortality than women whose tumors are detected by other methods.[41]

a. When interpreting their findings, the authors of the study discussed the possibility of lead-time bias accounting for the relatively better survival of women whose breast cancer was detected by means of screening mammography. Why might lead-time bias have been present?
b. To address this concern, the authors adjusted for tumor size and extent of disease when comparing survival across the groups of women with breast cancer defined by method of detection. What is the primary limitation of an analysis of this sort?

Answer 7.4

a. Mammography is a relatively more sensitive means of breast cancer detection: It can identify cancers earlier in their natural history than can self-exam or clinical exam. Therefore, even if there were no effective treatment for early breast cancer, the mean interval from diagnosis to death would be greatest in women whose tumors were identified via mammography, due to their added "lead-time" prior to the age they otherwise would have been diagnosed. In the absence of effective treatment, receipt of screening mammography would not have extended life, but rather it would have extended that portion of the same life span during which a woman was known to have breast cancer.
b. Screening has the potential to exert a favorable influence on the likelihood of cancer mortality only by identifying tumors that are relatively early in their natural history, in other words, small in size and limited in their spread. Adjusting for these characteristics would not allow whatever true benefit that screening has on survival to be evident.

Question 7.5 This question pertains to the following news article that appeared in the April 20, 2005, issue of the *Journal of the National Cancer Institute*:

> **Group Recommends Earlier Colorectal Cancer Screening for African Americans**
>
> The American College of Gastroenterology has recommended that physicians begin screening African Americans for colorectal cancer at age 45 rather than at age 50, the general recommendation made by several groups.
>
> In the publication of the recommendation, which appears in the March issue of the *American Journal of Gastroenterology*, the authors point out that African Americans have the highest incidence of colorectal cancer of any racial or ethnic group. In addition, they note that the mean age of presentation among African Americans is lower than whites.

The American College of Gastroenterology based its recommendation (regarding a racial difference in the age at which screening for colorectal cancer begins) on the high incidence and low mean age at presentation of colorectal cancer in African Americans. The latter is *not* a valid reason. Why?

Answer 7.5 It is possible that the basis for the relatively younger age distribution of African-American cases is the relatively younger age distribution of African Americans in general. The incidence rate of colorectal cancer among 45- to 49-year-olds—the relevant piece of data that would bear on the recommendation—could be identical between the races and the mean age at diagnosis would still be lower in African Americans than in whites.

Question 7.6 The following is excerpted from the abstract of an article on computer-aided detection of early breast cancer.[42]

> We determined the association between the use of computer-aided detection at mammography facilities and the performance of screening mammography from 1998 through 2002 at 43 facilities in three states. We had complete data for 222,135 women (a total of 429,345 mammograms), including 2351 women who received a diagnosis of breast cancer within 1 year after screening. We calculated the specificity, sensitivity, and positive predictive value of screening mammography with and without computer-aided detection.
>
> Diagnostic specificity decreased from 90.2% before implementation to 87.2% after implementation ($P < 0.001$). The increase in sensitivity from 80.4% before implementation of computer-aided detection to 84.0% after implementation was not significant ($P = 0.32$).
>
> The use of computer-aided detection is associated with reduced accuracy of interpretation of screening mammograms. The increased rate of biopsy with the use of computer-aided detection is not clearly associated with improved detection of invasive breast cancer.

The size of the decrease in specificity associated with computer-aided detection—90.2% – 87.2% = 3.0%—was slightly smaller than the size of the increase in sensitivity—84.0% – 80.4% = 3.6%. However, the p value for the difference in specificity (<0.001) was far smaller than that for the difference in sensitivity (0.32). What is the reason for the disparity between the size of two p values?

Answer 7.6 The calculation of the specificity is based on the women without breast cancer, whereas the calculation of the sensitivity is based on the women with breast cancer. The *p* value is heavily influenced by the number of subjects, and in this study the women without cancer outnumbered the women with cancer by nearly 100 to 1. Thus, the *p* value must be smaller for the difference in specificity than that for the difference in sensitivity.

Question 7.7 A randomized trial of fecal occult blood screening for colorectal cancer was conducted in Minnesota.[43]

a. During a follow-up period of 18 years, the incidence of colorectal cancer that was first diagnosed when metastatic (stage D) in 15,570 persons assigned to annual screening was only 53% that of 15,394 persons assigned to not be screened; for 15,587 persons assigned to be screened every 2 years, the corresponding figure was 68%. Are any other data needed to show that screening was successful?
b. The table below summarizes the mortality experience of the three study groups. In assessing the efficacy of fecal occult blood screening, should attention be focused primarily on total mortality or mortality from colorectal cancer?

	Study group		
	Annual screening	*Biennial screening*	*Control*
No. enrolled	15,570	15,587	15,394
Person-years of observation	240,325	240,163	237,420
Deaths from all causes			
No. of deaths	5,236	5,213	5,186
Cumulative mortality*	342	340	343
95% confidence interval (CI)	334–350	333–348	336–531
Deaths from colorectal cancer (CRC)			
No. of deaths	121	148	177
Cumulative mortality*	9.46	11.19	14.09
95% CI	7.75–11.17	9.39–12.99	12.01–16.17

* Per 1000.

Answer 7.7

a. The answer depends on the criterion for success. On average, persons in the groups assigned to receive screening did have their tumors detected at a relatively earlier stage than persons assigned not to be screened. However, unless treatment at this earlier stage is more efficacious than that given later, no lives will have been saved. Therefore, if one wishes to learn whether screening led to a reduction in mortality, it is necessary to compare mortality rates in the screened and unscreened participants.

b. If assessment of cause of death is believed to be largely unaffected by screening history and the impact of screening is believed to be confined to averting deaths from colorectal cancer, then the focus of the analysis should be on mortality from colorectal cancer. If there are potentially fatal complications of the screening process (e.g., perforated colon during colonoscopy that would follow a positive stool exam) or treatment (e.g., death during cancer surgery), deaths from these causes should be included as well.

   A comparison of all-cause mortality would be an insensitive means of measuring the potential benefit associated with screening: Even if annual screening prevented *every* death from colorectal cancer and had no bearing on any other cause of death, the all-cause mortality in the screened group would be decreased by only 3%

$$\left(\frac{5{,}186 - 177}{5{,}186} = 0.97\right)$$

Question 7.8 The following is excerpted from a manuscript describing the results of a cohort study of screening for stomach cancer (by means of photofluorography, a highly sensitive test) in Japan in relation to mortality from stomach cancer:

> The present study focused on 100,562 subjects who were ages 40–79 years at the time of a baseline survey, which asked for their screening experience during the past twelve months. Of these, 219 subjects with a history of stomach cancer prior to the time of the baseline survey were excluded, as were 8,386 subjects who did not provide information on their participation in stomach-cancer screening.
>
> Death rates from stomach cancer during a period following the baseline survey, adjusted for differences in age and other demographic characteristics, were compared between participants with and without a history of screening for this disease.

Do you believe this study is likely to provide a valid estimate of the ability of screening photofluorography to lead to a reduction in mortality from stomach cancer? If yes, why? If not, why not, and in which direction do you believe the bias will occur?

Answer 7.8 The design of this study does not permit an unbiased estimate of the potential impact of photofluorographic screening on mortality from stomach cancer. The problem stems from what has been referred to as "healthy screenee bias."[26] The "exposed" cohort in the present study has received a screening exam in the prior year that was negative: Had it been positive for cancer, such persons would have been removed from the cohort. The "unexposed" cohort has not had such persons identified or excluded. Thus, even if screening led to completely ineffective treatment, or to no treatment at all, the observed relative mortality from stomach cancer would be lower in the screened group.

Question 7.9 You find the following in an article in a medical journal:

> Pap smear screening may be less effective among black women than among white women. Laboratory-based evidence of Pap smear screening (i.e., a Pap smear performed in the absence of symptoms of cervical cancer) at least once during the past 5 years was found for 48% of the black population with invasive cervical cancer versus only 32% of the white population with invasive cervical cancer ($p < 0.05$).

Assuming that ascertainment of invasive cervical cancer was equally complete for the black and white populations and that management of the Pap smear abnormalities was similar for the two groups, what reason might there be for the observed result other than a differential effectiveness of Pap smear screening between the two races?

Answer 7.9 It is possible that in the underlying population from which the women with cervical cancer were drawn, a higher percentage of black women than white women had undergone Pap screening at least once in the past 5 years. What is needed in order to gauge the efficacy of this screening modality in the prevention of invasive cervical cancer is the proportion of women in the population at risk who had been screened. Suppose the following results would have been observed in a case-control study of invasive cervical cancer:

	Black women		*White women*	
	Invasive cancer	*Controls*	*Invasive cancer*	*Controls*
% screened	48	60	32	44
% not screened	52	40	68	56
Odds ratio	0.6		0.6	

In this situation, the relative impact of screening in reducing the incidence of invasive cervical cancer would have been identical in black and in white women.

Question 7.10 A randomized trial was performed to determine whether screening women with ovarian cancer in clinical remission for serum levels of a tumor marker every 3 months, followed by treatment of recurrences identified through this means, could reduce mortality. Women were recruited into the trial at the onset of the remission induced by their initial treatment. Those who later were found to have elevated levels of the marker were randomized to one of two approaches to management:

(a) Immediate chemotherapy; or
(b) No additional treatment until the recurrence became apparent for other reasons (typically, the development of symptoms), on average 5 months later. (Values for serum levels of the tumor marker in patients in group (b) were not provided to these women's physicians.)

In comparing mortality between women in groups (a) and (b), should rates be calculated from the time of randomization or from the time of recognition of the recurrence? Explain the reason for your answer.

Answer 7.10 The mortality rates in both groups should be based on follow-up beginning at the time of randomization, in other words, when the elevation in serum levels of the tumor marker was first apparent. Follow-up beginning at the time of recognition of the recurrence would give rise to lead-time bias: Only in group (a) would person-time be accrued during (on average) the first 5 months after tumor marker elevation was noted, and in this period of time mortality from ovarian cancer would be expected to be considerably smaller than in the period following symptom development.

(A randomized trial of the efficacy of screening for the tumor marker CA 125, designed in the manner described in this question, did indeed tabulate mortality rates in both groups from the time of randomization.[47])

Question 7.11 You are designing a case-control study to estimate the efficacy of PSA screening for prostate cancer. From the records of a large prepaid health-care plan, you are going to ascertain screening histories of men who died as a result of prostate cancer during 2007-2009. As a basis for comparison, you are considering two possible control groups (who would be individually matched to the fatal cases):

1. Members of the health-care plan who are demographically comparable to the fatal cases and also were diagnosed with prostate cancer at about the same time, but who were still alive at the time of the matched case's death
2. Demographically comparable members of the health plan who had not been diagnosed with prostate cancer as of the date of diagnosis of their matched fatal case

Which of the two groups above would provide the more valid result? Explain.

Answer 7.11 Option 2 conforms to the goal of control definition, in other words, a sample of the population from which the cases were derived. The problem with option 1 is that, even if no effective treatment were available for screen-detected prostate cancer (in which case the correct odds ratio should be one), there would be a higher proportion of screened men in controls than cases (assuming that PSA screening has some sensitivity for detecting prostate cancer).[48]

Question 7.12 A case-control study was conducted to assess the impact of cervical screening in reducing the incidence of invasive cervical cancer.[44] Because such screening can identify cervical precancerous lesions, it was hypothesized that a smaller proportion of cases than controls would have been screened at some point during the prior 4 years.

The women with cervical cancer had been diagnosed during 2000–2003; they were identified using records of a cancer registry serving the area in which the study was done. The receipt of screening during 1996–2003 was ascertained by means of a registry that enumerated screening exams in that area. Age-matched controls were chosen from women identified in the screening registry during 1996–2004. The investigators explained that "each case would still have the possibility of being matched to control(s) with no screening history prior to the case's diagnosis," since some controls would have received their first screen during 2004 and so would have been unscreened during 1996–2003.

What do you perceive to be the greatest threat to the validity of this study?

Answer 7.12 The purpose of a control group in a case-control study is to estimate the proportion of "exposed" persons in the population (who were at risk of the disease in question) from which the cases arose. In this instance, the investigators needed screening histories on a sample of women with a cervix who resided in the area from which the cases had been derived. It would be fortuitous if the sampling from the controls who were chosen by the investigators succeeded in accomplishing this. Almost certainly, the level of screening in the controls actually selected (from a screening registry) overstated that in the population at risk, leading to an observed odds ratio (approximately 0.1) that was falsely low.

Question 7.13 Hackam et al. ascertained prior use of antihypertensive and other medications among patients over 65 years of age hospitalized with a ruptured abdominal aortic aneurysm ("cases," n = 3,379) and those hospitalized with an unruptured aneurysm ("controls," n = 11,947).[45] In the course of reviewing electronic files of medical services that preceded the aneurysm diagnosis, the investigators incidentally noted receipt of abdominal imaging (ultrasound, CT, or MRI) in nearly all of the patients with an unruptured abdominal aortic aneurysm, but in only a small fraction of those with a ruptured aneurysm.

Screening by means of abdominal imaging has the potential to lead to identification of an aortic aneurysm, and in turn to surgical intervention that could avert a rupture. However, despite the large observed case-control difference in the prior receipt of screening by means of abdominal imaging, it would *not* be appropriate to use the data of Hackam et al. to support the hypothesis that screening could lead to a reduced risk of a ruptured abdominal aortic aneurysm. Why? What would be a better basis for comparison to the patients with a ruptured abdominal aortic aneurysm?

Answer 7.13 In order for a case-control study of screening efficacy to provide a valid result, the level of screening in the controls must reflect that of the population at risk from which the cases were derived. All of the controls in this study were known to have an abdominal aortic aneurysm, many, no doubt, as a result of screening, and so the level of screening in them would be expected to be atypically high. Thus, even if there truly were no beneficial impact of early recognition and treatment of abdominal aortic aneurysms on the occurrence of rupture, a (far) smaller proportion of persons with a ruptured aneurysm than a known unruptured one would have a history of screening.

If one were to design a case-control study to estimate the ability of aneurysm screening to reduce the occurrence of ruptured abdominal aortic aneurysm (this was not the goal of Hackam et al.), controls would need to be selected from an entirely different sampling frame. If (as in this study) the cases of ruptured aneurysm were all those over age 65 diagnosed in a specific geographic population, the controls ideally would be demographically similar members of that same population.

(Fortunately, there are data available from randomized trials from which we can gauge the efficacy of ultrasound screening for abdominal aortic aneurysm.[49])

Question 7.14 In order to gauge the efficacy of screening by means of a digital rectal exam (DRE) or measurement of levels of prostate-specific antigen (PSA) in serum against mortality from prostate cancer, a case-control study was conducted.[46] A comparison was made between receipt of either of these screening modalities in 74 men who died as a result of prostate cancer ("cases") and in a sample of year-of-birth matched men who resided in the same county as did the cases. Because of the concern that tests performed close to the time of the diagnosis of prostate cancer might have been prompted by symptoms, the primary analysis focused on screening tests performed during the 1 to 5 years prior to diagnosis of prostate cancer, and during the corresponding period of time in control men. Screening tests had been performed on 81.3% of controls during this interval, in contrast to just 60.8% of the fatal cases (odds ratio = 0.35, 95% confidence interval = 0.17–0.71).

Because of the means by which "exposure" was defined in the primary analysis, it is likely that a smaller proportion of cases than controls would have been screened, even had early identification of prostate cancer failed to lead to treatment that would favorably influence the risk of death from prostate cancer. Why is this?

Answer 7.14 For a screening modality that is sensitive (such as a serum PSA measurement for the presence of prostate cancer), excluding the 12 months prior to diagnosis from consideration will effectively exclude most positive tests. And, because positive tests are proportionally far more numerous in men with prostate cancer than in men in general, exclusion of the 12 months prior to diagnosis means that proportionally more cases than controls who had been screened will be classified as not having been screened. Therefore, even if screening failed to lead to any mortality reduction, this analytic approach would give rise to an odds ratio less than 1, suggesting efficacy against cancer mortality when in truth none had been present.

A valid result in case-control study of screening requires accurate assessment of receipt of screening in cases and controls during the whole of the period prior to diagnosis or symptoms (whichever comes first) that correspond to the period of potential detectability.[50]

REFERENCES

1. Prescott E, Hippe M, Schnohr P, et al. Smoking and the risk of myocardial infarction in women and men: longitudinal population study. *BMJ* 1998;316:1043–1046.
2. Kendal WS. Suicide and cancer: a gender-comparative study. *Ann Oncol* 2007;18:381–387.
3. Andriole GL, Crawford ED, Grubb RL, et al. Mortality results from a randomized prostate-cancer screening trial. *N Engl J Med* 2009; 360:1310–1309.
4. Carroll-Pankhurst C, Mortimer EA. Sudden infant death syndrome, bedsharing, parental weight, and age at death. *Pediatrics* 2001; 107: 530–536.
5. Kumar AS, Benz CC, Shim V, et al. Estrogen receptor-negative breast cancer is less likely to arise among lipophilic statin users. *Cancer Epidemiol Biomarkers Prev* 2008; 17:1028–1033.
6. Mosley BS, Cleves MA, Siega-Riz AM, et al. Neural tube defects and maternal folate intake among pregnancies conceived after folic acid fortification in the United States. *Am J Epidemiol* 2009;169:9–17.
7. Wong YN, Mitra N, Hudes G, et al. Survival associated with treatment vs observation of localized prostate cancer in elderly men. *JAMA* 2006;296:2683–2693.

8. Galea S, Blaney S, Nandi A, et al. Explaining racial disparities in incidence of and survival from out-of-hospital cardiac arrest. *Am J Epidemiol* 2007;166:534–543.
9. Stromberg, B, Dahlquist G, Ericson A, et al. Neurological sequelae in children born after in-vitro fertilisation: a population-based study. *Lancet* 2002;359:461–465.
10. Koepsell TD, Weiss NS. Epidemiologic Methods. Oxford, New York, 2003.
11. Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. *JAMA* 1998;280: 605–613.
12. Cardo DM, Culver DH, Ciesielski CA, et al. A case-control study of seroconversion in health care workers. Centers for Disease Control and Prevention Needlestick Surveillance Group. *N Engl J Med* 1997;337:1485–1490.
13. Steinmaus C, Yuan Y, Bates MN, et al. Case-control study of bladder cancer and drinking water arsenic in the western United States. *Am J Epidemiol* 2003;158:1193–1201.
14. Furness S, Connor J, Robinson E, et al. Car colour and risk of car crash injury: population based case control study. *BMJ* 2003;327:1455–1456.
15. Ray JG, Meier C, Vermeulen MJ, et al. Association of neural tube defects and folic acid food fortification in Canada. *Lancet* 2002;360:2047–2048.
16. Lemaitre RN, Siscovick DS, Raghunathan TE, et al. Leisure-time physical activity and the risk of primary cardiac arrest. *Arch Int Med* 1999;159:686–690.
17. Jick SS, Hagberg KW, Kaye JA, et al. Postmenopausal estrogen-containing hormone therapy and the risk of breast cancer. *Obstet Gynecol* 2009;113:74–80.
18. Skjaerven R, Wilcox AJ, Lie RT. The interval between pregnancies and the risk of preeclampsia. *N Engl J Med* 2002;346:33–38.
19. Villeneuve PJ, Holowaty EJ, Brisson J, et al. Mortality among Canadian women with cosmetic breast implants. *Am J Epidemiol* 2006;164:334–341.
20. Nichol KL, Baken L, Wuorenma J, et al. The health and economic benefits associated with pneumococcal vaccination

of elderly persons with chronic lung disease. *Arch Intern Med* 1999; 159: 2437–2442.
21. Leonard RC, Kreckmann KH, Sakr CJ, et al. Retrospective cohort mortality study of workers in a polymer production plant including a reference population of regional workers. *Ann Epidemiol* 2008;18:15–22.
22. Stattin P, Holmberg, E, Johansson, JE, et al. Outcomes in localized prostate cancer: National Prostate Cancer Register of Sweden followup study. *J Nat Cancer Inst* 2010;102:950–958.
23. Anderson JP, Ross JA, Folsom AR. Anthropometric variables, physical activity, and incidence of ovarian cancer. *Cancer* 2004; 100:1515–1521.
24. Beral V, Million Women Study Collaborators. Breast cancer and hormone-replacement therapy in the Million Women Study. *Lancet* 2003;362:419–427.
25. Youakim S. Risk of cancer among firefighters: a quantitative review of selected malignancies. *Arch Environ Occup Health* 2006;61:223–231.
26. Weiss NS, Rossing MA. Healthy screenee bias in epidemiologic studies of cancer incidence. *Epidemiology* 1996;7:319–322.
27. Caughey AB, Nicholson JM, Cheng YW, et al. Induction of labor and cesarean delivery by gestational age. *Am J Obstet Gynecol* 2006;195:700–705.
28. Chiazze L, Ference LD, Wolf PH: Mortality among automobile assembly workers: I. Spray painters. *J Occup Med* 1984;22: 520–526.
29. Rowe B, Milner R, Johnson C, et al. The association of alcohol and night driving with fatal snowmobile injury: a case-control study. *Ann. Emerg Med* 1994;24:842–848.
30. Heffelfinger JD, Heckbert SR, Psaty BM, et al. Influenza vaccination and risk of incident myocardial infarction. *Hum Vaccin* 2006;2:161–166.
31. Spirtas R, Heineman EF, Bernstein L, et al. Malignant mesothelioma: attributable risk of asbestos exposure. *Occup Environ Med* 1994; 51:804–811.
32. Rosenberg L, Rao RS, Palmer JR, et al. Transitional cell cancer of the urinary tract and renal cell cancer in relation to acetaminophen use. *Cancer Causes Control* 1998;9:83–88.

33. Handke M, Harloff A, Olschewski M, et al. Patent foramen ovale and cryptogenic stroke in older patients. *N Engl J Med*. 2007;351:2262–2268.
34. Hessel PA, Teta MJ, Goodman M, et al. Mesothelioma among brake mechanics: an expanded analysis of a case-control study. *Risk Anal* 2004;24:547–552.
35. Grant AM, Avenell A, Campbell MK, et al. Oral vitamin D3 and calcium for secondary prevention of low-trauma fractures in elderly people: a randomised placebo-controlled trial. *Lancet* 2005; 365:1621–1628.
36. Glynn RJ, Danielson E, Fonseca FA, et al. A randomized trial of rosuvastatin in the prevention of venous thromboembolism. *N Engl J Med* 2009;360:1851–1861.
37. John EM, Koo J, Horn-Ross PL. Lifetime physical activity and risk of endometrial cancer. *Cancer Epidemiol Biomarkers Prev* 2010;19:1276–1283.
38. Donnell D, Baeten JM, Kiarie J, et al. Heterosexual HIV-1 transmission after initiation of antiretroviral therapy: a prospective cohort analysis. *Lancet* 2010;375:2092–2098.
39. Treggiari MM, Weiss NS. Occupational asbestos exposure and the incidence of non-Hodgkin lymphoma of the gastrointestinal tract: an ecologic study. *Ann Epidemiol* 2004;14:168–171.
40. Yuan JM, Gao YT, Ong CN, et al. Prediagnostic level of serum retinol in relation to reduced risk of hepatocellular carcinoma. *J Natl Cancer Inst* 2006;98:482–490.
41. McPherson CP, Swenson KK, Jolitz G, et al. Survival of women ages 40-49 years with breast carcinoma according to method of detection. *Cancer* 1997;79:1923–1932.
42. Fenton JJ, Taplin SH, Carney PA, et al. Influence of computer-aided detection on performance of screening mammography. *N Engl J Med* 2007; 356:1399–1409.
43. Mandel JS, Church TR, Ederer F, et al. Colorectal cancer mortality: effectiveness of biennial screening for fecal occult blood. *J Natl Cancer Inst* 1999;91:434–437.
44. Yang B, Morrell S, Zuo Y, et al. A case-control study of the protective benefit of cervical screening against invasive cervical cancer in NSW women. *Cancer Causes Control* 2008;19:569–576.

45. Hackam DG, Thiruchelvam D, Redelmeier DA. Angiotensin-converting enzyme inhibitors and aortic rupture: a population-based case-control study. *Lancet* 2006; 368:659–665.
46. Bergstralh EJ, Roberts RO, Farmer SA, et al. Population-based case-control study of PSA and DRE screening on prostate cancer mortality. *Urology* 2007;70:936–941.
47. Rustin GJ, van der Burg ME, Griffin CL, et al. Early versus delayed treatment of relapsed ovarian cancer. *Lancet* 2010;376: 1155–1163.
48. Weiss NS. Control definition in case-control studies of the efficacy of screening and diagnostic testing. *Am J Epidemiol* 1983;118:457–460.
49. Thompson SG, Ashton HA, Gao L, Scott RA. Screening men for abdominal aortic aneurysm: 10 year mortality and cost effectiveness results from the randomised Multicentre Aneurysm Screening Study. *BMJ* 2009;338:b2307.
50. Weiss NS, Dhillon PK, Etzioni RS. Case-control studies of the efficacy of cancer screening. Overcoming bias from nonrandom patterns of screening. *Epidemiology* 2004;15:409–413.

This page intentionally left blank

INDEX