

Epidemiology and Biostatistics

Practice Problem Workbook

Bryan Kestenbaum



Springer

Epidemiology and Biostatistics

Bryan Kestenbaum

Epidemiology and Biostatistics

Practice Problem Workbook

 Springer

Bryan Kestenbaum, MD, MS
Division of Nephrology
Department of Medicine
University of Washington
Seattle, WA
USA

ISBN 978-3-319-97432-3 ISBN 978-3-319-97433-0 (eBook)
<https://doi.org/10.1007/978-3-319-97433-0>

Library of Congress Control Number: 2018953296

© Springer Nature Switzerland AG 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This workbook was created as a companion to the second edition of the textbook, *Epidemiology and Biostatistics: An Introduction to Clinical Research*. The questions and answers in the book are designed to encourage hands-on application of the concepts taught in the textbook with special emphasis on common areas of difficulty. Many of the questions are intended to parallel material covered by the United States Medical License Examination. However, the broad intention of this workbook is to enhance interpretation of real-world medical and public health studies using practical examples that cover fundamental aspects of study design, sources of bias and error, screening and diagnostic testing, and statistical analyses.

The problems in this workbook intend to capture the unique perspective of learning Epidemiology and Biostatistics for the first time. Essential to the creation of this book were the many thoughtful and probing questions of the students.

Seattle, WA, USA

Bryan Kestenbaum, MD, MS

Contents

1	Measures of Disease Frequency	1
2	Population, Exposure, and Outcome.	5
3	Case Reports and Case Series	7
4	Cross-Sectional Studies.	9
5	Cohort Studies.	13
6	Case-Control Studies	17
7	Randomized Trials	21
8	Misclassification	27
9	Confounding	31
10	Effect Modification	37
11	Screening and Diagnosis.	41
12	Summary Measures in Statistics	49
13	Statistical Inference.	53
14	Hypothesis Tests in Practice.	57
15	Linear Regression	61
16	Log-Link and Logistic Regression.	65
17	Survival Analysis.	69
18	Practice Problem Workbook Solutions.	73

Chapter 1

Measures of Disease Frequency



Measures of disease frequency quantify the burden and development of disease in populations. Two common measures of disease frequency are prevalence and incidence. Prevalence provides a snapshot of the amount of disease that is present at a specific point or period in time. Prevalence data are useful for raising awareness of disease and allocating health resources but may be insufficient for establishing temporal relationships between potential risk factors and disease. Incidence describes the development of new disease over time. In a given population, the prevalence of a disease is proportional to the incidence and the disease duration.

Researchers examine the association of beta-carotene supplement use with diabetes. They identify 20 patients who report regular use of beta-carotene from a local clinic and 20 patients from the same clinic who do not report use of beta-carotene. The researchers determine diabetes status at the start of the study and then annually over 5 years of follow-up by querying the patients’ electronic medical records. Raw study data are presented below.

Patient number	Beta-carotene use	Follow-up time (months)	Diabetes present at the start of study	New diabetes diagnosed during follow-up	Reason for leaving study
1	Yes	44	No	Yes	New diabetes diagnosis
2	Yes	60	No	No	Study ended
3	Yes	32	No	No	Dropout
4	Yes	40	Yes	No	Lost to follow-up
5	Yes	60	No	No	Study ended
6	Yes	60	No	No	Study ended

Patient number	Beta-carotene use	Follow-up time (months)	Diabetes present at the start of study	New diabetes diagnosed during follow-up	Reason for leaving study
7	Yes	18	No	Yes	New diabetes diagnosis
8	Yes	60	Yes	No	Study ended
9	Yes	32	No	Yes	New diabetes diagnosis
10	Yes	60	No	No	Study ended
11	Yes	60	No	No	Study ended
12	Yes	50	No	No	Dropout
13	Yes	26	Yes	No	Lost to follow-up
14	Yes	28	No	No	Lost to follow-up
15	Yes	34	No	No	Dropout
16	Yes	8	No	Yes	New diabetes diagnosis
17	Yes	60	No	No	Study ended
18	Yes	60	No	No	Study ended
19	Yes	40	No	Yes	New diabetes diagnosis
20	Yes	60	Yes	No	Study ended
21	No	20	No	Yes	New diabetes diagnosis
22	No	48	No	No	Lost to follow-up
23	No	44	No	Yes	New diabetes diagnosis
24	No	42	No	No	Dropout
25	No	6	No	No	Dropout
26	No	60	No	No	Study ended
27	No	60	No	No	Study ended
28	No	28	No	No	Lost to follow-up
29	No	60	No	No	Dropout
30	No	12	No	No	Lost to follow-up
31	No	22	No	No	Dropout
32	No	30	No	Yes	New diabetes diagnosis
33	No	60	No	No	Lost to follow-up
34	No	22	No	Yes	New diabetes diagnosis
35	No	28	No	No	Lost to follow-up
36	No	60	Yes	No	Study ended
37	No	60	Yes	No	Study ended
38	No	24	No	Yes	Dropout
39	No	20	No	No	Dropout
40	No	50	No	No	Lost to follow-up

1. What is the prevalence of diabetes in this study population at the start of the study (baseline)?
2. What is the incidence proportion of diabetes in this study population during follow-up?
3. What is the incidence proportion of diabetes among patients who use beta-carotene?
4. What is the incidence rate of diabetes among patients who use beta-carotene?
5. What is the incidence rate of diabetes among patients who do not use beta-carotene?
6. Triple antiviral therapy has dramatically improved survival among patients with human immunodeficiency virus (HIV) disease. If the incidence of HIV were to remain constant, what is the expected impact of widespread triple antiviral therapy on the prevalence of HIV in the population?
 - A. Increase
 - B. Decrease
 - C. Stay the same
7. The incidence of a disease is five times greater in men compared with women, yet there is no difference in disease prevalence by sex. What is the best explanation for this finding?
 - A. Men receive more intensive medical care for the disease.
 - B. The mortality rate is greater among women.
 - C. The disease is less aggressive among women.
 - D. Women are older than men when they are diagnosed with the disease.

Anecdotal evidence suggests that anxiety disorder may contribute to the irritable bowel syndrome (IBS), a condition characterized by nausea, alternating constipation and diarrhea, and no identifiable gastrointestinal pathology. Researchers administer an online questionnaire regarding IBS symptoms to 10,000 people who have an established diagnosis of anxiety disorder in the United States, Canada, and Mexico. They administer the same questionnaire to another 10,000 people from the same countries who do not have a diagnosis of anxiety disorder. Their findings are tabulated below.

	IBS symptoms	No IBS symptoms
Anxiety disorder ($N = 10,000$)	4000	6000
No anxiety disorder ($N = 10,000$)	1000	9000

8. Which of the following is true?
 - A. The incidence proportion of IBS symptoms among people with anxiety disorder is 40%.
 - B. The incidence density of IBS symptoms among people with anxiety disorder is 40%.

- C. The prevalence of IBS symptoms among people with anxiety disorder is 40%.
- D. The relative risk of IBS symptoms among people with anxiety disorder is 40%.
9. Which of the following represents a reasonable next step based on the study findings?
- A. Provide access to educational materials about IBS to patients who have anxiety disorder.
- B. Increase the use of antianxiety medications to prevent symptoms of IBS.
- C. Submit case reports describing patients who have a diagnosis of anxiety disorder with concomitant IBS symptoms.
- D. Perform laboratory work to investigate mechanisms by which anxiety disorder might stimulate gastrointestinal nerve transmission.

A company consults you to assess the risk of carpal tunnel syndrome among its employees. You interview 1000 employees to inquire about their work status and whether they developed carpal tunnel syndrome since working for the company. Results are shown below.

	Number of employees	New instances of carpal tunnel syndrome
Full-time employees	800	25
Part-time employees	200	6
Incidence proportion among all employees = $(31 \text{ cases}/1000 \text{ people}) \times 100\% = 3.1\%$		

The CEO of the company is concerned that this amount of carpal tunnel syndrome is considerably higher than that of a competitor company.

10. Which denominator would permit the most accurate comparison of the incidence of carpal tunnel syndrome between the two companies?
- A. Number of employees
- B. Number of sedentary hours
- C. Number of person hours
- D. Number of carpal tunnel syndrome cases

Chapter 2

Population, Exposure, and Outcome



A study population refers to all people who enter a research study, regardless of whether they are exposed, are treated, develop the outcome of interest, or drop out of the study before completion. The exposure and outcome of a study depend on the proposed study question. The exposure refers to any characteristic that may explain or predict the presence of a study outcome. The outcome refers to the characteristic that is being predicted.

A study investigates whether neonatal hyperbilirubinemia is associated with the risk of future language delay in children. Researchers identify 100 infants who have neonatal hyperbilirubinemia and a comparison group of 100 infants who do not have this condition. They next determine rates of language delay after 3 years.

11. What are the study population, exposure, and outcome of this study?

A study explores characteristics that may influence the use of creatine, a dietary supplement, among high school students. Researchers interview 1200 students from 5 metropolitan high schools to inquire about creatine use, dietary habits, physical activity, and smoking. The researchers estimate caloric intake from the reported dietary data. The study finds that greater caloric intake is associated with a higher likelihood of creatine use among boys, but not girls.

12. What are the study population, exposure, and outcome of this study?

A study evaluates whether surgical experience is associated with the risk of bile leak after laparoscopic cholecystectomy. Researchers identify 800 general surgeons from a large health care organization. They use the medical information system to ascertain the number of previous laparoscopic cholecystectomy procedures and the number of bile leaks for each surgeon. The study finds that the incidence of postoperative bile leak is lower among surgeons who perform more laparoscopic cholecystectomies.

13. What are the study population, exposure, and outcome of this study?

14. Which of the following exclusion criteria would be *least* suitable for the study of the laparoscopic cholecystectomy described above?

- A. Exclusion of patients with a previous history of bile leak
 - B. Exclusion of patients with a previous history of bile duct disease, which can increase the risk of postoperative bile leak
 - C. Exclusion of surgeons with missing data regarding the number of previous laparoscopic cholecystectomy procedures
 - D. Exclusion of surgeons who have performed fewer than 20 previous laparoscopic cholecystectomy procedures
15. Which of the following would promote *internal validity* of the study of laparoscopic cholecystectomy described above?
- A. Evaluating the association of surgical experience with postoperative mortality
 - B. Adding data from other healthcare organizations across geographic regions
 - C. Assessing other types of laparoscopic surgical procedures and outcomes
 - D. Performing medical chart review to confirm the presence of postoperative bile leaks that occurred during the study
16. Which of the following would promote *external validity* of the study of laparoscopic cholecystectomy described above?
- A. Excluding patients who have a previous history of bile leak
 - B. Adding data from other healthcare organizations across geographic regions
 - C. Calculating incidence rates of postoperative bile leak using person-time data
 - D. Performing medical chart review to confirm the presence of postoperative bile leaks that occurred during the study

Chapter 3

Case Reports and Case Series



Case reports and case series describe the experience of people who have a specific disease or condition. These studies can be useful for raising awareness of new diseases and can generate hypotheses regarding possible causes. However, case reports and case series have inherent limitations that hinder inference of causal relationships: lack of a suitable denominator to calculate incidence, absence of comparison groups, small sample size, and ambiguous external validity.

The following criteria can be used to infer causal relationships in research studies:

- Strength of association.
- Biologic plausibility.
- Association varies predictably across levels of exposure (dose-response).
- Randomized evidence.
- Temporal relationship between exposure and outcome.

Complement factor H is a regulatory component of the alternate complement pathway that helps protect host cells from immune-mediated damage. Experimental and human genetic studies suggest that complement factor H may play a role in preventing age-related macular degeneration. A study measures circulating complement factor H levels in 30 patients who have a confirmed diagnosis of macular degeneration. The study finds that complement factor H levels are abnormally low in 9 (30%) of these patients.

17. How many criteria are met by this study regarding a possible causal impact of complement factor H levels on age-related macular degeneration?
- A. 0
 - B. 1
 - C. 2
 - D. 3
 - E. 4

Antipsychotic medications can prolong ventricular repolarization, which may increase the risk of certain cardiac arrhythmias. A study reports six cases of torsades de pointes, a life-threatening arrhythmia, among people who were taking conventional antipsychotic medications. Review of medical records reveals that none of the study individuals had a previous history of heart disease or arrhythmia and that the arrhythmias resolved following the administration of magnesium.

18. How many criteria are met by this study regarding a possible causal impact of antipsychotic medication use on torsades de pointes?
- A. 0
 - B. 1
 - C. 2
 - D. 3
 - E. 4

A newspaper article reports the following information:

“Defective tires have been linked with several recent sport utility vehicle (SUV) crashes. We investigated tires obtained from 30 recent SUV accidents. We found that 24 of these tires were manufactured at a single plant in Cincinnati, Ohio. An investigation is underway to determine whether there are significant problems in the assembly line of the Cincinnati plant.”

19. What is the most important problem with concluding that the Cincinnati plant is particularly at fault from these data?
- A. Employees at the Cincinnati plant may have different characteristics than employees at other manufacturing plants.
 - B. Training procedures at the Cincinnati plant may differ from those at other plants.
 - C. The investigators evaluated only tires from sport utility vehicles.
 - D. Investigators failed to look for more subtle signs of tire damage in other vehicles.
 - E. Lack of a denominator.
20. Following the introduction of a new immunosuppressant medication for lung transplantation, a study describes the occurrence of unusual opportunistic infections among eight recent users of the medication. Which of the following would be the most appropriate next step for investigating this problem?
- A. Randomized trial of the new immunosuppressant medication with opportunistic infections as the primary outcome
 - B. Removal of the new immunosuppressant medication from the market
 - C. Prospective cohort study to evaluate long-term rates of mortality among users and nonusers of the immunosuppressant medication
 - D. Determining incidence rates of opportunistic infections among users and nonusers of the new immunosuppressant medication

Chapter 4

Cross-Sectional Studies



Cross-sectional studies are defined by measurement of the exposure and the outcome of a study at the same time. There is no follow-up time in cross-sectional studies. Consequently, cross-sectional study data are useful for determining the relative prevalence of a disease or condition but are typically inadequate for discerning temporal relationships unless there is strong plausibility for one of the directions of association.

Questions 21–24 refer to the article by Drybe et al.:

Burnout and Serious Thoughts of Dropping Out of Medical School: A Multi-Institutional Study. *Acad Med.* 2010;85(1):94–102.

21. What is the relative prevalence (or prevalence ratio) of having severe thoughts of dropping out of medical school, comparing students who have children with students who do not have children, in the cross-sectional component of this study?
 - A. 2.80
 - B. 1.80
 - C. 0.55
 - D. 0.50
 - E. Cannot determine from the data in the paper
22. What is the overall prevalence of burnout among students in the cross-sectional component of this study?
 - A. 8.8%
 - B. 18.6%
 - C. 22.8%
 - D. 49.4%
 - E. Cannot determine from the data in the paper

23. What is the incidence rate of having severe thoughts of dropping out of medical school in this study?

- A. 65%
- B. 0.46 per 1000 student-years
- C. 0.86 per 1000 student-years
- D. 1.32 per 1000 student-years
- E. Cannot determine from the data in the paper

The following criteria can be used to infer causal relationships:

- Strength of association.
- Biologic plausibility.
- Association varies predictably across levels of exposure (dose-response).
- Randomized evidence.
- Temporal relationship between exposure and outcome.

24. How many criteria regarding a possible causal impact of burnout on serious thoughts of dropping out of medical school are met by this study?

- A. 0
- B. 1
- C. 2
- D. 3
- E. 4

A study investigates characteristics that may be associated with low circulating vitamin D levels. Investigators recruit cohorts of Caucasian and African American adults from five US communities, determine race and sex by self-report, and measure height, weight, and serum 25-hydroxyvitamin D levels, the accepted storage form of vitamin D.

First, the investigators evaluate the association of race with serum vitamin D levels:

	<i>Mean serum 25-hydroxyvitamin D concentration (ng/mL)</i>
Caucasian (<i>N</i> = 1500)	32.5
African American (<i>N</i> = 1500)	20.2

25. Do these findings permit conclusion of a temporal association between race and serum 25-hydroxyvitmain D levels?

- A. Yes
- B. No

Next, the investigators evaluate the associations of obesity with serum vitamin D levels:

	<i>Mean serum 25-hydroxyvitamin D concentration (ng/mL)</i>
Obese (<i>N</i> = 900)	21.7
Nonobese (<i>N</i> = 2100)	28.4

26. Do these findings permit conclusion of a temporal association between obesity and serum 25-hydroxyvitmain D levels?
- A. Yes
 - B. No
27. Which of the following would be the most appropriate next step for studying the potential association between serum vitamin D levels and obesity?
- A. Conducting a randomized clinical trial of vitamin D supplementation among obese individuals with diabetes as the primary outcome
 - B. Enacting policies to promote greater vitamin D use in the general population
 - C. Enacting policies to increase testing for vitamin D deficiency
 - D. Evaluating the association between serum 25-hydroxyvitamin D levels and the prevalence of obesity in different countries
 - E. Determining incidence rates of obesity according to baseline serum levels of 25-hydroxyvitamin D

Chapter 5

Cohort Studies



Cohort studies are observational studies that determine the incidence of a disease or condition over time. The primary advantage of cohort studies over cross-sectional studies is the ability to separate potential risk factors from the occurrence of disease over time to assess temporal relationships. Cohort study data can be used to calculate measures of risk, including relative risk, attributable risk, and population attributable risk. Like other types of observational studies, the primary limitation of cohort studies is the possibility that characteristics other than the exposure of interest could impact the outcome of the study (confounding).

Researchers conduct a study to compare the risk of hypoglycemia (low blood sugar) among patients with diabetes who initiate long-acting versus short-acting insulin therapy. They recruit 15 patients who recently initiated insulin treatment and have no previous history of hypoglycemic episodes. Participants are followed for up to 2 years to assess occurrences of hypoglycemia. Results are shown below.

Subject number	Insulin type	Follow-up time (years)	Hypoglycemic episode
1	Long-acting	1.3	No
2	Long-acting	1.1	Yes
3	Long-acting	0.8	No
4	Long-acting	1.6	No
5	Long-acting	1.4	Yes
6	Long-acting	1.9	No
7	Long-acting	1.7	No
8	Short-acting	0.6	No
9	Short-acting	1.9	No
10	Short-acting	1.1	No
11	Short-acting	0.8	Yes
12	Short-acting	0.4	No

Subject number	Insulin type	Follow-up time (years)	Hypoglycemic episode
13	Short-acting	1.3	Yes
14	Short-acting	0.7	No
15	Short-acting	1.4	No

28. Which of the following best describes the design of this study?

- A. Cohort study
- B. Cross-sectional study
- C. Case report
- D. Case series
- E. Randomized trial

29. What is the relative risk of a hypoglycemic reaction, comparing long-acting insulin with short-acting insulin in this study?

- A. 1.00
- B. 0.92
- C. 0.84
- D. 192 cases per 1000 person-years
- E. 151 cases per 1000 person-years

A study examines the association of air pollution with asthma among children living in urban populations. Investigators recruit 1300 children ages 4–8 years old from pediatric clinics in 5 large cities. Study children are free of asthma or reactive airway disease at the start of the study. The investigators obtain baseline air pollution data from monitors located near each child's residence and conduct annual follow-up exams to determine new instances of asthma. They categorize exposure to fine particulate air pollution as low versus high based on a cut point of 15 $\mu\text{g}/\text{m}^3$. The study results are shown below.

Air pollution level	Asthma	No asthma	Person-time (years)
Low	200	800	3300
High	100	200	780

Assume for a moment that the children in this study are reasonably representative of children in the selected cities and that the observed differences in asthma incidence are solely due to air pollution.

30. Given these assumptions, how much additional asthma can be attributed to air pollution among children ages 4–8 in these cities who are exposed to high levels of air pollution?

- A. 1.3 cases per 100 person-years
- B. 6.7 cases per 100 person-years
- C. 7.4 cases per 100 person-years
- D. 9.5 cases per 100 person-years
- E. 12.8 cases per 100 person-years

31. Given these assumptions, how much additional asthma can be attributed to air pollution among the population of children ages 4–8 in these cities?
- A. 1.3 cases per 100 person-years
 - B. 6.7 cases per 100 person-years
 - C. 7.4 cases per 100 person-years
 - D. 9.5 cases per 100 person-years
 - E. 12.8 cases per 100 person-years.

Questions 32–34 refer to the article by Wang et al.:

Risk of Death In Elderly Users of Conventional Vs. Atypical Antipsychotic Medications. *N Engl J Med.* 2005;353(22):2335–41.

Note: The term “hazard ratio” appears several times throughout the article. For the purposes of these questions, consider a hazard ratio to be the same as relative risk.

32. All of the following represent advantages of the selected study design for examining whether atypical antipsychotic medications increase the risk of death *except one*:
- A. Ability to assess the risks and benefits of atypical antipsychotic medication use in a large population of older adults
 - B. Ability to assess the risks and benefits of atypical antipsychotic medication use among vulnerable people who may be excluded from randomized trials
 - C. Ability to assess the risks and benefits of atypical antipsychotic medication use in real-world settings
 - D. Ability to clearly distinguish the causal impact of atypical antipsychotic medication use on mortality from the characteristics of people who tend to use these medications
 - E. Ability to investigate the possibility of a dose-response association between atypical antipsychotic medication use and mortality
33. What is the attributable risk of death within 180 days comparing conventional antipsychotic medication use with atypical antipsychotic medication use?
- A. 1.51
 - B. 1.37
 - C. 3.3%
 - D. 4.5%
 - E. 4.5 cases per 100 person-years

The following criteria can be used to infer a causal relationship between the use of conventional antipsychotic medications and all-cause mortality:

- Strength of association.
- Biologic plausibility.
- Association varies predictably across levels of exposure (dose-response).
- Randomized evidence.
- Temporal relationship between exposure and outcome.

Note: Of the many associations presented in the study, consider the adjusted relative risk of 1.37 in Table 2 to represent the “primary” association between the type of antipsychotic medication use and mortality.

34. How many of the causal criteria are met by this study?

- A. 0
- B. 1
- C. 2
- D. 3
- E. 4

Chapter 6

Case-Control Studies



Case-control studies are a specialized type of observational study design ideally suited for evaluating rare diseases and those with a long latency period. Case-control studies begin by targeting people who have and do not have a disease or condition of interest and then work backward to determine associations with previous exposures. Due to the manner in which participants are selected for these studies, case-control data alone cannot be used to directly calculate the incidence of disease or incidence-based measures of risk. However, it is possible to estimate relative risk from case-control study data using the odds ratio.

Researchers investigate a potential link between hepatitis B vaccination and the occurrence of multiple sclerosis. They identify 100 patients from a network of neurology clinics who have received a diagnosis of multiple sclerosis and 100 healthy control individuals from primary care clinics in the same geographic region. The researchers query medical records to determine the proportion of people who had previously received the hepatitis B vaccine. Results are shown below.

	Number who previously received hepatitis B vaccine
100 patients with multiple sclerosis	12
100 healthy individuals	13

35. Which of the following is true?
- A. Given that multiple sclerosis is relatively uncommon in the population, the incidence proportion of multiple sclerosis among hepatitis B-vaccinated individuals is approximately 34%.
 - B. The attributable risk of multiple sclerosis associated with hepatitis B vaccination is approximately 1%.

- C. The odds ratio of multiple sclerosis comparing hepatitis B-vaccinated to non-hepatitis B-vaccinated individuals is 0.91.
- D. The study design clearly distinguishes the causal impact of the hepatitis B vaccine on multiple sclerosis from the characteristics of people who received this vaccine.

36. Which of the following is true?

- A. The internal validity of these study findings would be further enhanced by confirmation of the diagnosis of multiple sclerosis using magnetic resonance imaging.
- B. A more suitable control population would be healthcare workers who are required to receive hepatitis B vaccination.
- C. A more suitable control population would be uninsured adults who are less likely to receive vaccinations.
- D. Recall bias is an important potential problem in this study.

Questions 37–40 refer to the article by Travis et al.:

Bladder and Kidney Cancer Following Cyclophosphamide Therapy for Non-Hodgkin's Lymphoma. *J Natl Cancer Inst.* 1995;87(7):524–30.

37. What is the crude (unadjusted) odds ratio of bladder cancer associated with receiving cyclophosphamide without radiation, compared to receiving radiation without cyclophosphamide?
- A. 0.8
 - B. 2.5
 - C. 1.2
 - D. 0.4
 - E. Cannot be determined using the study data
38. What is the incidence of bladder cancer among lymphoma patients who received 50 or more grams of cyclophosphamide?
- A. 71%
 - B. 29%
 - C. 14.5 cases per 100 person-years
 - D. 6.3 cases per 100 person-years
 - E. Cannot be determined from the study data
39. Each of the following characteristics of the study represents a desirable methodological attribute of case-control studies *except one*:
- A. Evaluation of patients who had a confirmed diagnosis of non-Hodgkin lymphoma
 - B. Use of chemotherapy and radiation data that were collected before the development of secondary cancers to eliminate recall bias

- C. Use of pathology records to confirm the diagnosis of secondary bladder cancers
 - D. Selection of cases and controls from the same underlying population
 - E. Study of a rare disease to permit interpretation of odds ratios as relative risks
40. Which of the following is true?
- A. Cyclophosphamide plus radiation is associated with a similar risk of bladder cancer compared to cyclophosphamide without radiation.
 - B. Cumulative cyclophosphamide dosage is more important than duration in terms of bladder cancer risk.
 - C. Higher radiation dose is associated with a greater risk of bladder cancer.
 - D. Men have a greater odds of bladder cancer compared to women in this study.

Chapter 7

Randomized Trials



Randomized trials are interventional studies that administer specific treatments or control procedures to study participants. The primary advantage of the randomized design is the ability to separate the causal impact of treatments from the characteristics of people who receive these treatments. Characteristics that impact the internal validity of randomized studies include the randomization procedure, choice of comparator, blinding, concealment, adherence, accuracy of the study measurements, and the analytic plan. The results of randomized trials may have limited external validity due to preferential inclusion of relatively healthy participants and careful monitoring procedures.

A clinical trial tests whether scheduled exercise can prevent the development of diabetes among obese adults. Researchers establish a multi-site consortium that enrolls 3400 participants from clinics in 8 US cities. Inclusion criteria are a body mass index (BMI) $>30 \text{ kg/m}^2$ and no previous history of diabetes. Exclusion criteria include any physical or medical condition that would preclude regular exercise or a previous history of heart failure. Characteristics of enrolled participants are 72% female, mean age of 44 years, and mean BMI of 34 kg/m^2 .

Participants are randomized in a 1:1 ratio to receive either a scheduled exercise program or no such treatment. Participants in the exercise group receive a gym membership within close proximity of their residence and a suggested workout routine prescribing 150 min of aerobic activity per week. Study personnel contact participants in the exercise group every 3 months to encourage compliance with the program. Participants in the no-treatment group receive educational materials describing the importance of exercise at the start of the study.

All participants complete annual study visits to assess the development of diabetes, which is defined by a fasting blood glucose level $>126 \text{ mg/dL}$ or the new use of a medication for diabetes. The researchers are prevented from knowing which treatment was administered throughout trial. Study results over a median of 5 years of follow-up are presented below.

Group	Number of people	Number of diabetes cases	Diabetes incidence per 100 people	Diabetes incidence per 1000 person-years
Assigned to exercise	1700	55	3.2	6.5
Assigned to no treatment	1700	74	4.4	8.7

41. Which of the following criticisms of the trial is most valid?
- A. There is concern for bias due to the relatively high proportion of women who were enrolled in the trial.
 - B. There is concern for bias due to differences in age between the exercise and no-treatment groups.
 - C. There is concern for bias due to differences in fasting blood sugars between the exercise and no-treatment groups.
 - D. There is concern for bias because the researchers failed to exclude people with a family history of diabetes, a strong risk factor for the study outcome.
 - E. Inadequate blinding may have disproportionately impacted the health behaviors of participants in the exercise and no treatment groups.
42. Which of the following is most likely to jeopardize the *internal validity* of the trial?
- A. Exclusion of people with a history of heart failure
 - B. Use of a surrogate endpoint
 - C. Lack of restricted randomization procedures
 - D. The possibility that other characteristics of people assigned to the exercise program, and not exercise itself, may have caused the observed difference in diabetes rates
 - E. The possibility of nonadherence with the scheduled exercise program
43. Which of the following is most likely to jeopardize the *external validity* of the trial?
- A. Frequent contacts from study personnel to encourage compliance with exercise.
 - B. Participants who were assigned to no treatment may have exercised on their own.
 - C. Participants who were assigned to exercise may have engaged in other healthy behaviors outside of the trial.
 - D. Inadequate blinding.
 - E. Potential errors in measuring the study outcome.

44. What is the relative risk of diabetes, comparing participants who were assigned to the exercise program with participants who were assigned to no treatment?
- A. 8.70
 - B. 2.20
 - C. 1.33
 - D. 0.86
 - E. 0.75
45. What is the attributable risk of diabetes, comparing participants who were assigned to the exercise program with participants who were assigned to no treatment?
- A. -8.70 events per 1000 person-years
 - B. -2.20 events per 1000 person-years
 - C. -1.33 events per 1000 person-years
 - D. -0.86 events per 1000 person-years
 - E. -0.75 events per 1000 person-years
46. Approximately how many people similar to those in the trial would need to be treated with the exercise program to prevent one instance of diabetes over a median of 5 years of follow-up (note: use the diabetes incidences per 100 people for this calculation)?
- A. 12
 - B. 24
 - C. 37
 - D. 68
 - E. 83
47. If the same exercise program were administered to a community of sedentary people who are otherwise similar to the trial participants, how much of a reduction in the rate of diabetes would be expected in that community?
- A. 8.70 events per 1000 person-years
 - B. 2.20 events per 1000 person-years
 - C. 1.33 events per 1000 person-years
 - D. 0.86 events per 1000 person-years
 - E. 0.75 events per 1000 person-years

The investigators are concerned about the possibility of low adherence in the exercise group. A secondary analysis of the trial data reveals that only 1105 (65%) of participants assigned to the exercise program maintained compliance with this program during the trial. Moreover, among the 1700 participants assigned to no treatment, 260 reported initiating a regular exercise program on their own during the trial period.

48. Which of the following analytic strategies would best preserve the initial similarity in participant characteristics that was created by randomization?
- A. Compare diabetes outcomes among the 1105 participants who complied with the exercise program to the 1700 participants who were assigned to no treatment.
 - B. Compare diabetes outcomes among the 1105 participants who complied with the exercise program to the 1440 participants who did not initiate an exercise program.
 - C. Compare diabetes outcomes among the 1700 participants who were assigned to the exercise program to the 1700 participants who were assigned to no treatment.
 - D. Compare diabetes outcomes among the 1700 participants who were assigned to the exercise program to the 1440 participants who did not initiate an exercise program on their own.

Returning to the original trial data, the investigators next explore whether the exercise program may have been particularly beneficial among people who were sedentary at the start of trial. They define sedentary behavior by self-report of at least 8 h per day of inactivity. Previous experimental studies have demonstrated that transition from a sedentary lifestyle to even modest levels of activity produces particularly large increases in glucose uptake by skeletal muscle tissue and improvement in insulin sensitivity. Trial data stratified by the presence of sedentary behavior at baseline are presented below.

Group	Number of people	Number of diabetes cases	Diabetes incidence per 1000 person-years
Self-reported sedentary behavior			
Assigned to exercise	600	24	8.0
Assigned to no treatment	600	31	10.3
Self-reported no sedentary behavior			
Assigned to exercise	1100	31	5.6
Assigned to no treatment	1100	43	7.8

49. Do these findings support a differential impact of the exercise program by the presence of sedentary behavior at baseline?
- A. Yes
 - B. No
 - C. Insufficient information to decide

Researchers plan to evaluate a new oral immune-modulating therapy for locally advanced breast cancer. They conceive a randomized trial to test the new treatment in women who have hormone receptor-negative, human epidermal growth factor receptor 2 (HER2)-negative cancers. Currently accepted treatment for this condition includes radiotherapy and intravenous chemotherapy.

50. Which of the following comparison groups would best preserve blinding in a randomized trial of the new agent while maintaining equipoise?
- A. No treatment
 - B. A placebo
 - C. Radiotherapy and intravenous chemotherapy plus a placebo
 - D. Radiotherapy and intravenous chemotherapy
 - E. Delayed treatment

A randomized trial compared acupuncture with physical therapy for the treatment of chronic low back pain. Researchers identified potential participants with low back pain from primary care practices located in Portland, Oregon. They excluded people who had any a previous history of cancer or vertebral fracture. Participants were randomized in a 1:1 ratio to receive either a prescribed acupuncture program or regular physical therapy sessions twice weekly. The primary outcome of the trial was the change in back pain after 8 weeks, determined by a standardized back pain questionnaire.

51. Which one of the following characteristics impacts the *external validity* of this trial?
- A. Lack of blinding
 - B. Subjective outcome measure
 - C. Recruitment from practices in Portland, Oregon
 - D. Potential nonadherence with acupuncture
 - E. Potential nonadherence with physical therapy

Questions 52–55 refer to the article by Cummings et al.:

Denosumab for Prevention of Fractures in Postmenopausal Women with Osteoporosis. *N Engl J Med.* 2009;361(8):756–65.

52. Which of the following represents the greatest threat to the *internal validity* of the trial?
- A. Restriction of the study population to women who had T-scores between -2.5 and -4.0 .
 - B. Exclusion for long-term bisphosphonate use.
 - C. Use of a surrogate endpoint.
 - D. Potential age differences between the denosumab and placebo groups.
 - E. None of the above – the study findings are internally valid.
53. Based on the raw numbers of hip fractures that occurred in the trial and the total numbers of participants in the denosumab and placebo groups listed in Table 1, how many similar postmenopausal women with osteoporosis would need to be treated with denosumab to prevent one hip fracture within 36 months?
- A. 17
 - B. 167
 - C. 233
 - D. 500
 - E. 644

54. Which of the following represents the most appropriate response to the greater number of cellulitis cases observed among denosumab-treated participants?
- A. Nothing – this is likely a statistical anomaly.
 - B. Nothing – the number of cellulitis cases is numerically small.
 - C. Post-approval surveillance.
 - D. Follow-up clinical trial of denosumab among patients who are at high risk for cellulitis.
 - E. Restriction of the label indication to patients who are at low risk for cellulitis.
55. In a hypothetical secondary analysis, investigators use medication claims data to discover that 20% of the trial participants who were assigned to placebo initiated denosumab treatment outside of the trial. Given this new finding, how should the original study data be analyzed?
- A. No change – continue to compare participants who were initially assigned to denosumab to those who were initially assigned to placebo.
 - B. Exclude participants from the placebo group if they initiated denosumab treatment outside of the trial.
 - C. Discontinue follow-up time in the placebo group at the time participants first initiated denosumab outside of the trial.
 - D. Consider participants in the placebo group to be taking a placebo until the time of first denosumab use and then consider them to be taking denosumab thereafter.
 - E. Re-weight the relative risks by the proportion of outside denosumab use.

Chapter 8

Misclassification



Misclassification refers to the false characterization of a study characteristic due to measurement error. Information regarding the procedures used to measure the study data helps to infer whether misclassification is likely to have occurred and the suspected type of misclassification. Non-differential misclassification arises from non-systematic error in measuring the study data and in most instances leads to observing a relative risk that is closer to 1.0 than that obtained under ideal measurements (bias toward the null). Differential misclassification arises from systematic error in measuring the study data, the impact of which depends on the specific pattern of measurement error that has occurred.

Researchers conduct a case-control study to examine the association of paint exposure with pulmonary fibrosis, a serious disease that typically presents with shortness of breath and a nonproductive cough. The researchers identify 30 case individuals who have received a diagnosis of pulmonary fibrosis, confirmed by high-resolution computed tomography, and a comparison group of 90 healthy control individuals who are free of pulmonary symptoms. The researchers conduct in-person interviews with the case and control individuals to inquire about previous exposures to latex and oil-based paint products.

56. Which type of misclassification would be most important to consider in this study?
- A. Nonselective misclassification of the exposure
 - B. Selective misclassification of the exposure
 - C. Nonselective misclassification of the outcome
 - D. Selective misclassification of the outcome

A related case-control study is conducted among workers from a large national construction company. Researchers identify 30 employees who have received a diagnosis of pulmonary fibrosis, confirmed by high-resolution computed tomography, and a group of 90 control employees from the same company who are free of pulmo-

nary symptoms. The researchers query the company's computerized job assignment records to ascertain previous exposures to latex and oil-based paint products.

57. Which type of misclassification would be most important to consider in this study?
- A. Nonselective misclassification of the exposure
 - B. Selective misclassification of the exposure
 - C. Nonselective misclassification of the outcome
 - D. Selective misclassification of the outcome

Researchers use data from a large health maintenance organization to examine the association of aspirin use with all-cause mortality. They identify a cohort of 10,000 people who were regularly using aspirin as of January 1, 2000, and a second cohort of 40,000 people who were not using aspirin on the same date. The researchers link enrollee records with the National Death Index, a centralized database of mortality data, to determine vital status through 2017. The study finds a small, but statistically insignificant association of aspirin use with a lower risk of all-cause mortality: relative risk 0.94 (95% confidence interval 0.81, 1.07).

58. How does the relative risk observed in this study likely compare with the relative risk that would have been obtained if the study data were measured perfectly?
- A. The observed relative risk is likely to be higher than the relative risk obtained under idealized study measurements.
 - B. The observed relative risk is likely to be lower than the relative risk obtained under idealized study measurements.
 - C. The observed relative risk is likely to be the same as the relative risk obtained under idealized study measurements.

A follow-up study is conducted to assess the validity of the aspirin use data. The researchers link enrollee records to electronic pharmacy databases to determine the consistency of aspirin use over time. They find that 30% of enrollees who were originally classified as aspirin users subsequently discontinued aspirin use over the next 5 years. In contrast, nearly all enrollees who were originally classified as non-aspirin users at baseline remained non-aspirin users throughout the study.

59. Based on this information, how does the relative risk observed in this study likely compare with the relative risk that would have been obtained if the data were measured perfectly?
- A. The observed relative risk is likely to be higher than the relative risk obtained under idealized study measurements.
 - B. The observed relative risk is likely to be lower than the relative risk obtained under idealized study measurements.
 - C. The observed relative risk is likely to be the same as the relative risk obtained under idealized study measurements.

A cohort study evaluates the association of apolipoprotein B (ApoB) levels with incident myocardial infarction. Researchers measure plasma ApoB levels using an automated laboratory platform and define an “elevated” ApoB level by the highest 20% of measured values. The study finds that elevated ApoB levels are associated with a twofold greater incidence of myocardial infarction over 5 years of follow-up.

To check the validity of the automated laboratory platform, the researchers compare ApoB levels measured on this platform with ApoB levels measured using a gold standard laboratory method on the same blood sample. They find that the automated platform consistently returns values that are 50% lower than those of the reference laboratory.

60. Based on this information, how does the relative risk observed in this study likely compare to the relative risk that would have been obtained if the data were measured perfectly?
- A. The observed relative risk is likely to be higher than the relative risk obtained under idealized study measurements.
 - B. The observed relative risk is likely to be lower than the relative risk obtained under idealized study measurements.
 - C. The observed relative risk is likely to be the same as the relative risk obtained under idealized study measurements.

Chapter 9

Confounding



In observational studies, associations between potential risk factors and disease may or may not represent causal relationships. Confounding is a type of bias that occurs when characteristics other than the exposure of interest distort the observed association of the exposure with disease. A confounding characteristic is defined as a factor that is associated with the exposure, associated with the outcome, and not suspected to reside on the causal pathway of association. Strategies to control for confounding include restriction, stratification plus adjustment, matching, and regression. The presence of confounding is suspected when the size of the association of interest changes meaningfully after adjustment by one of these methods.

HIV disease may increase susceptibility to other viral infections. A cohort study investigates the association of HIV with the occurrence of cytomegalovirus (CMV) infection, a common herpesvirus. Researchers screen infectious disease clinics to identify a cohort of 400 HIV-positive patients who are seronegative for CMV (indicating no previous exposure). The researchers then identify a comparison cohort of 400 people without HIV disease from primary care clinics who are also CMV seronegative.

Study personnel conduct annual testing to assess new CMV infections, defined by the development of antibodies to the virus. The study data are presented in the following tables (Tables 9.1 and 9.2).

61. Which of the following characteristics is most likely to have confounded the observed association of HIV disease with incident CMV infection?
- A. Age
 - B. Sex
 - C. Body mass index
 - D. Intravenous drug use
 - E. CD4 lymphocyte count

Table 9.1 Baseline characteristics of the study participants

	HIV positive	HIV negative
Age (years)	47.3	47.1
African American (%)	37.3	18.9
Male (%)	54.0	52.9
Body mass index (kg/m ²)	23.2	27.9
Intravenous drug use (%)	35.4	4.1
Mean CD4 lymphocyte count (cells/mm ³)	187	1440

Table 9.2 Associations of study characteristics with incident CMV infection

	Unadjusted relative risk of CMV infection
HIV disease	4.05
Age (per 10-year higher)	2.92
African American (compared to Caucasian)	1.01
Male (compared to female)	2.05
Body mass index (per 5 kg/m ² higher)	1.03
Intravenous drug use (yes versus no)	1.86
CD4 lymphocyte count (per 100 cells/mm ³ increase)	2.70

62. Which description best describes how race relates to the observed association between HIV disease and CMV infection?
- A. Associated with exposure, associated with outcome, not on causal pathway
 - B. Associated with exposure, associated with outcome, on causal pathway
 - C. Associated with exposure, not associated with outcome, not on causal pathway
 - D. Associated with exposure, not associated with outcome, on causal pathway
 - E. Not associated with exposure, associated with outcome, not on causal pathway

The director of a general internal medicine clinic is concerned about possible harm caused by the overtreatment of hypertension. She compiles data from all hypertensive patients who received treatment in the clinic over the past 10 years, including average treated systolic blood pressures and mortality status over follow-up. Findings are shown below.

Treated systolic blood pressure (mmHg)	Number of patients	Mortality rate (per 1000 person-years)
<90	45	14.6
90–110	97	7.5
110–130	212	6.1
130–150	397	6.9
>150	108	9.9

63. Which one of the following statements is FALSE?

- A. These data demonstrate that lowering systolic blood pressure <90 mmHg increases the risk of death.
- B. The higher mortality rate observed among patients who had treated systolic blood pressures <90 mmHg may be confounded by age.
- C. The higher mortality rate observed among patients who had treated systolic blood pressures <90 mmHg may be confounded by low body weight.
- D. The higher mortality rate observed among patients who had treated systolic blood pressures <90 mmHg may be confounded by comorbidity.
- E. The higher mortality rate observed among patients who had treated systolic blood pressures <90 mmHg may have been caused by decreased cerebral blood flow.

There is growing enthusiasm for green tea as a protective agent against cancer, possibly due to antioxidant properties. An observational study reports that the regular consumption of green tea is associated with a 30% lower incidence of pancreatic cancer compared with no green tea consumption: unadjusted relative risk 0.7.

People who regularly consumed green tea in the study were, on average, younger and more likely to exercise regularly, consumed more dietary fiber, and had greater circulating concentrations of antioxidants compared with non-green tea drinkers. Moreover, when subjected to provocative tests of vascular function, green tea drinkers in the study exhibited a greater vasodilatory response, suggesting lower oxidative stress. Characteristics that were associated with a higher risk of pancreatic cancer in this study included older age, lower circulating concentrations of antioxidants, and a reduced vasodilatory response.

64. Which of the following characteristics is most likely to have confounded the observed association between green tea consumption and pancreatic cancer in this study?

- A. Age
- B. Regular exercise
- C. Dietary fiber
- D. Circulating antioxidants
- E. Vasodilatory response

The researchers conducting the study decide to adjust for circulating concentrations of antioxidants. After adjustment, they find regular green tea consumption to be associated with only a 10% lower incidence of pancreatic cancer compared with no green tea consumption: unadjusted relative risk 0.9.

65. Which of the following is true?

- A. Circulating concentrations of antioxidants were likely to have been confounding the unadjusted association between green tea consumption and pancreatic cancer.
- B. The researchers should match green tea drinkers to non-green tea drinkers by circulating concentrations of antioxidants.

- C. The researchers should perform stratified analyses by circulating concentrations of antioxidants.
- D. The investigators should restrict their analyses to participants who had low circulating concentrations of antioxidants.
- E. Circulating concentrations of antioxidants are likely to reside on the causal pathway of association between green tea consumption and pancreatic cancer.

The researchers conducting the study decide to adjust for sex by determining the association of green tea consumption with pancreatic cancer separately among men and women. Results are shown below.

	Unadjusted relative risk of pancreatic cancer comparing regular green tea drinkers with non-green tea drinkers
All study participants	0.70
Men only	0.85
Women only	0.85

66. Which of the following is true?

- A. The stratified data suggest that sex is confounding the unadjusted association between green tea consumption and pancreatic cancer.
- B. The stratified data suggest that sex is likely to reside on the causal pathway between green tea consumption and pancreatic cancer.
- C. The stratified data suggest that other factors are likely to be confounding the unadjusted association between green tea consumption and pancreatic cancer.
- D. The use of stratification is not appropriate.
- E. The stratified results are not possible.

Large placebo-controlled randomized trials have reported that long-acting calcium channel blockers reduce the risk of stroke among people with hypertension. In an observational study, researchers obtain clinical data from enrollees in five large healthcare systems. They observe stroke rates of 5.0 per 1000 person-years among enrollees who are treated with long-acting calcium channel blockers compared with 3.0 per 1000 person-years among enrollees who are not treated with these medications.

67. Which of the following is most likely to account for the discrepancy between the findings reported from the randomized trials and the results observed in the observational study?

- A. Lack of blinding
- B. Lack of a denominator
- C. Lack of a plausible mechanism of association
- D. Confounding by indication
- E. Errors in classifying long-acting calcium channel blocker use

68. Which of the following strategies would be most helpful for investigating the potential causal impact of long-acting calcium channel blocker use on stroke risk in the observational study?
- A. Match long-acting calcium channel blocker users to nonusers by sex.
 - B. Restrict the study population to enrollees who have hypertension.
 - C. Select long-acting calcium channel blocker users who have consistently received this medication for at least 3 years.
 - D. Compare the risk of stroke between long-acting calcium channel blocker users and aspirin users.
 - E. Use a random procedure to select users and nonusers of long-acting calcium channel blockers in the observational study.
69. Methods to control for confounding include all of the following *except*:
- A. Regression
 - B. *P*-values and 95% confidence intervals
 - C. Stratification
 - D. Matching
 - E. Restriction

Questions 70–71 refer to the article by Schaer et al.:

Risk for Incident Atrial Fibrillation in Patients Who Receive Antihypertensive Drugs: A Nested Case-Control Study. *Ann Intern Med.* 2010;152(2):78–84.

70. Which of the following is true regarding age as a possible confounder in this study?
- A. Age is not associated with the outcome; however, data regarding the association of age with the exposure is needed to decide whether age is a confounder.
 - B. Age is not associated with the outcome; however, data regarding the causal pathway of association is needed to decide whether age is a confounder.
 - C. Age is not associated with the exposure; however, data regarding the association of age with the outcome is needed to decide whether age is a confounder.
 - D. Age is likely to reside on the causal pathway of association.
 - E. Age is not likely to be a confounder in this study.

The study demonstrates an association of total ACE inhibitor use with a lower odds of atrial fibrillation. Specifically, the unadjusted odds ratio for total ACE inhibitor use is 0.81 and the adjusted odds ratio is 0.79 (Table 3, row 3).

71. Which of the following is true?
- A. The unadjusted association is likely to be confounded by smoking.
 - B. The unadjusted association is likely to be confounded by body mass index.

- C. The adjusted association represents the true causal impact of total ACE inhibitor use on the incidence of atrial fibrillation.
- D. There is a 21% lower odds of incident atrial fibrillation comparing total ACE inhibitor users with nonusers holding all of the characteristics listed in the footnote of Table 3 constant.

Chapter 10

Effect Modification



The impact of an exposure or treatment may differ across people. Effect modification occurs when the size of an association between an exposure and outcome differs according to another characteristic. A true differential impact of an exposure or treatment on the disease process is suspected when the difference in the size of an association across subgroups is large, the subgroups contain sufficient numbers of people and outcomes for comparison, there is plausibility for the observed difference, and the difference is replicated in other studies. Non-overlapping confidence intervals and a low p -value for interaction provide statistical evidence for the presence of effect modification.

A pharmaceutical company sponsors a clinical trial of a new medication for treating major depressive disorder. Investigators randomly assign 5000 people who are experiencing a first major depressive episode to receive either the new medication or a selective serotonin reuptake inhibitor (SSRI), the clinical standard of care for major depression. Among the recruited participants, the mean age is 37 years, 60% are female, and 22% also meet criteria for anxiety disorder. The outcome of the study is the persistence of depressive symptoms after 6 months of treatment. Results are shown below.

Trial results stratified by the presence of baseline anxiety disorder.

	Relative risk of persistent depression symptoms comparing the new medication to SSRIs (95% confidence interval)
All participants	0.70 (0.61, 0.79)
Anxiety disorder at baseline	0.38 (0.14, 0.62)
No anxiety disorder at baseline	0.79 (0.65, 0.93)

72. How does the presence of anxiety disorder relate to the observed association between receipt of the new antidepressant medication and persistent depressive symptoms?
- | Confounder | Effect modifier |
|---|-----------------|
| A. No | Yes |
| B. No | Yes |
| C. Yes | No |
| D. Yes | No |
| E. Cannot determine from the information that is provided | |

A cohort study evaluates the association of hepatitis B infection with the development of hepatic cirrhosis. Researchers identify 800 patients who have serologic evidence of active hepatitis B infection and a comparison group of 1800 healthy individuals who do not have evidence of past or present hepatitis B infection. Baseline characteristics of the study population are presented below.

	Hepatitis B positive	Hepatitis B negative
Mean age (years)	43.5	47.2
Male	68%	41%
Smoking	22%	14%
Heavy alcohol use	18%	9%
Chronic kidney disease	8%	11%

Participants are followed over 10 years to assess the occurrence of cirrhosis. The study finds that hepatitis B infection is associated with a 2.5 times higher incidence of cirrhosis:
Unadjusted relative risk 2.5; 95% confidence interval (2.3, 2.7)
Stratified study results are shown below.

Association of hepatitis B infection with cirrhosis among subgroups (relative risks and 95% confidence intervals shown)	
Men	Women
2.9 (2.1, 3.7)	2.3 (1.6, 3.0)
Heavy alcohol use	No heavy alcohol use
6.8 (5.4, 8.2)	1.9 (1.6, 2.2)

73. Which of the following is true?
- A. These data suggest that sex is confounding the observed association between hepatitis B infection and cirrhosis.
 - B. These data suggest that sex is modifying the observed association between hepatitis B infection and cirrhosis.
 - C. The sex-adjusted relative risk of cirrhosis associated with hepatitis B infection is between 2.3 and 2.9.

- D. These data suggest that heavy alcohol use is confounding the observed association between hepatitis B infection and cirrhosis.
- E. These data suggest that heavy alcohol use lies on the causal pathway between hepatitis B infection and cirrhosis.

A new vaccine designed to prevent malaria is deployed in four hypothetical countries. Disease incidences among vaccinated and non-vaccinated residents in each country are presented below.

	Incidence of malaria in non-vaccinated individuals (per 1000 at risk)	Incidence of malaria in vaccinated individuals (per 1000 at risk)
Country 1	100	40
Country 2	200	130
Country 3	300	230
Country 4	400	310

74. Consider the possibility that funds are available to continue the vaccination program in only one country. Deploying the new vaccine to which country would prevent the most cases of malaria?
- A. Country 1
 - B. Country 2
 - C. Country 3
 - D. Country 4

Acute kidney injury (AKI) can result from exposure to medications that reduce blood flow to the kidneys. Two common medication classes that can decrease kidney blood flow are nonsteroidal anti-inflammatory drugs (NSAIDs) and angiotensin receptor blockers (ARBs). A case-control study evaluates the association of these medications with the incidence of AKI. Researchers identify 100 case patients who recently developed AKI and 300 control patients who have normal kidney function. The researchers then ascertain NSAID and ARB use within the previous 6 months.

Odds ratio of AKI comparing NSAID use to nonuse: 1.6 (95% confidence interval 1.25, 1.95)

Odds ratio of AKI comparing ARB use to nonuse: 2.2 (95% confidence interval 1.65, 2.75)

75. Which of the following procedures would be most helpful for judging whether there is a possible synergistic effect of NSAID and ARB use on the risk of AKI?
- A. Evaluating the differential size of the odds ratios for AKI associated with NSAID and ARB use
 - B. Evaluating the 95% confidence intervals for the odds ratios of AKI associated with NSAID use and ARB use
 - C. Calculating *p*-values for the odds ratios of AKI associated with NSAID use and ARB use
 - D. Evaluating the association of NSAID use with AKI separately among ARB users and nonusers

Questions 76–78 refer to the article by Muzaale et al.:
Risk of End-Stage Renal Disease Following Live Kidney Donation. *JAMA*. 2014;311(6):579–86.

76. Based on the data in the paper, how does age relate to the association between live kidney donation and the development of end-stage renal disease?

- | | |
|---------------------|------------------|
| Confounder | Effect modifier |
| A. No | No |
| B. Yes | No |
| C. No | Yes |
| D. Cannot determine | Yes |
| E. No | Cannot determine |
77. Which of the following is true regarding the association of live kidney donation with ESRD?
- A. The attributable risk is greater among Black donors compared with Hispanic donors.
 - B. The relative risk is greater among Black donors compared with Hispanic donors.
 - C. The attributable risk is greater among donors who are ≥ 50 years old compared with donors who are < 50 years old.
 - D. The relative risk is greater among donors who are ≥ 50 years old compared with donors who are < 50 years old.

A hypothetical follow-up study obtains physical activity data for the live kidney donors and matched non-donors in the study. The association of live kidney donation with end-stage renal disease according to categories of low versus high levels of physical activity is shown below.

	Attributable risk of end-stage renal disease comparing live kidney donors to healthy non-donors
All study individuals	+26.9 per 10,000
Low physical activity	+12.3 per 10,000
High physical activity	+13.8 per 10,000

78. Based on these data, how do physical activity levels relate to the association between live kidney donation and end-stage renal disease?

- | | |
|---|-----------------|
| Confounder | Effect modifier |
| A. No | No |
| B. No | Yes |
| C. Yes | No |
| D. Yes | Yes |
| E. Cannot determine from the information that is provided | |

Chapter 11

Screening and Diagnosis



Screening and diagnostic tests can detect diseases at early stages to promote potentially beneficial treatments. The interpretation of screening and diagnostic test results depends on the disease process of interest, the qualities of the test, and the context of testing. Screening tests seek to identify diseases before they have caused recognizable signs or symptoms. Diagnostic tests seek to verify the presence of diseases for which signs and symptoms are already apparent. The predictive values of any test are determined by the test validity (sensitivity and specificity) and the pretest probability of the disease. Observational studies that compare outcomes of people who undergo screening versus those who do not undergo such procedures require careful interpretation to ensure that bias has not occurred.

79. Which of the following conditions would be most suitable for screening?

- A. Influenza
- B. Mad cow disease
- C. Melanoma
- D. Eczema
- E. Insomnia

A screening test for lead toxicity in children has a reported sensitivity of 95% and specificity of 80%. The test is administered to 1000 children from a neighborhood in which the prevalence of lead toxicity is 2%.

80. How many children who do *not* have actual lead toxicity would be expected to test positive (false positive test)?

- A. 1
- B. 19
- C. 196
- D. 784
- E. 215

81. The same test for lead toxicity is administered to 1000 children from a neighborhood in which the prevalence of lead toxicity is 50%. If a child from this population tests negative, what is the probability that he/she truly does *not* have lead toxicity?

- A. 10%
- B. 17%
- C. 50%
- D. 80%
- E. 94%

The sensitivity and specificity of the HIV antibody test are 99.7% and 98.5%, respectively. There are approximately one million people in the United States who have HIV disease.

82. If the entire United States population of 300 million people were screened with the HIV antibody test, approximately how many would be expected to have a false positive test result?

- A. 1 million
- B. 2 million
- C. 4.5 million
- D. 299 million
- E. 294.5 million

83. What is the positive predictive value of the HIV test when applied to the entire US population?

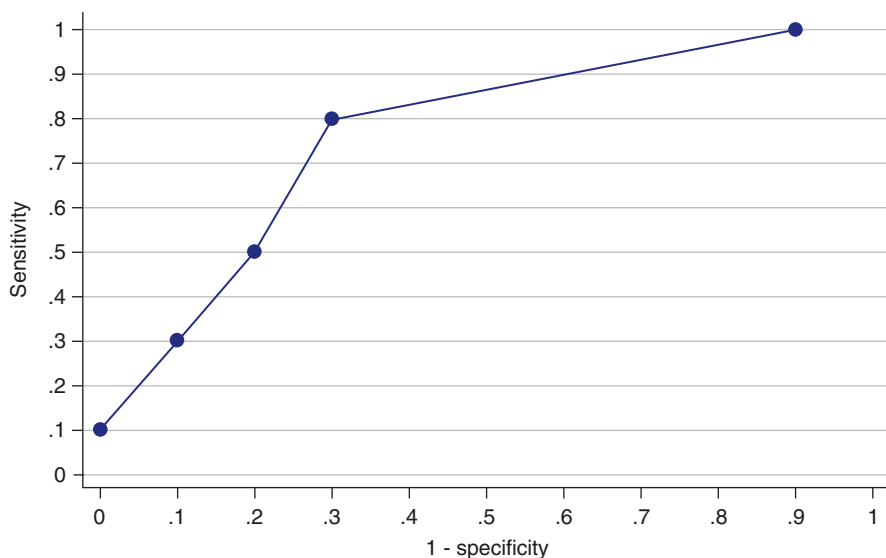
- A. 11%
- B. 18%
- C. 41%
- D. 76%
- E. 97%

A man presents to his primary care physician complaining of low-grade fevers, diarrhea, and a 15 lb weight loss over the past 3 months. He has used intravenous drugs for 10 years. Physical examination reveals enlarged cervical and femoral lymph nodes. Based on the history and physical examination findings, you estimate that this patient has an approximate 30% pretest probability of having HIV disease and order the HIV antibody test.

84. If the test comes back positive, what is the probability that this patient has HIV disease?

- A. 11%
- B. 18%
- C. 41%
- D. 76%
- E. 97%

A five-point test is created for diagnosing rheumatoid arthritis. Patients receive one point for each of the following characteristics: morning stiffness, joint swelling, erosive joint disease, a positive serum rheumatoid factor, and an elevated serum C-reactive protein level. The lowest possible score for the new test is 0 (no characteristics present) and the highest possible score is 5 (all characteristics present). The receiver operating characteristic curve below presents the performance of the test in relation to a gold standard diagnosis of rheumatoid arthritis.



85. Which of the following cut points for a positive test would provide the best overall balance between sensitivity and specificity?
- A. Defining a positive test by a score of 1 or more points
 - B. Defining a positive test by a score of 2 or more points
 - C. Defining a positive test by a score of 3 or more points
 - D. Defining a positive test by a score of 4 or more points
 - E. Defining a positive test by a score of 5 points
86. Which of the following definitions of a positive test would most reliably rule out rheumatoid arthritis among people who have a negative test?
- A. Defining a positive test by a score of 1 or more points
 - B. Defining a positive test by a score of 2 or more points
 - C. Defining a positive test by a score of 3 or more points
 - D. Defining a positive test by a score of 4 or more points
 - E. Defining a positive test by a score of 5 points

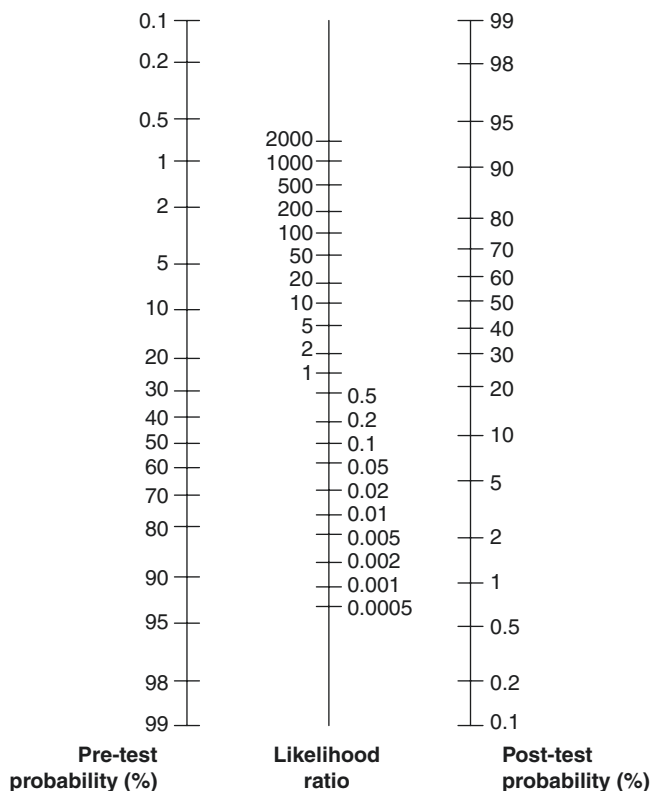
Researchers develop a new serologic test for cancer. They administer the test to 100 patients who have biopsy confirmed pancreatic cancer; 99 of these patients test positive. They next administer the test to 100 individuals who are known to be free of pancreatic cancer; only two test positive.

Excited by these findings, the researchers decide to evaluate the utility of the new test in clinical practice. They administer the test to 10,000 apparently healthy adults from a large, multispecialty practice in which the prevalence of pancreatic cancer is 1%. Individuals who test positive undergo subsequent computed tomography scanning and, if necessary, a pancreas biopsy.

87. Among the 10,000 tested individuals, how many total computed tomography scans are expected to result from this testing program?
- A. 45
 - B. 77
 - C. 98
 - D. 198
 - E. 297

The rapid antigen detection test for streptococcal pharyngitis has a positive and negative likelihood ratio of 35 and 0.3, respectively. You are called to the urgent care center to evaluate a child who has sudden onset of sore throat, exudative pharyngitis, and a fever of 103 °F. You estimate the pretest probability of streptococcal pharyngitis in this child to be 60%.

88. If the rapid antigen test is positive, what is the probability that this child has actual streptococcal pharyngitis? Use the likelihood ratio nomogram on the following page.
- A. Approximately 98%
 - B. Approximately 85%
 - C. Approximately 70%
 - D. Approximately 52%
 - E. Approximately 40%



Researchers develop a new rapid diagnostic test for rhinovirus, a common cause of upper respiratory infections. They administer the test to 100 consecutive people who present to an urgent care clinic with a runny nose, fever, and cough. In parallel, they determine the presence of actual rhinovirus infection using real-time polymerase chain reaction (PCR), the gold standard method for viral detection, in the same 100 individuals. The researchers compare the results of the new diagnostic test with the presence of actual rhinovirus infection determined by PCR.

89. Which characteristic of the new test does this study evaluate?

- A. Reliability
- B. Predictive value
- C. Validity
- D. Preclinical disease
- E. Utility

Researchers develop a new test to screen for chronic kidney disease (CKD). They administer the test to 5000 apparently healthy people from six communities. A central laboratory conducts all of the testing. A total of 450 of the tested individuals have a positive test.

The researchers select a comparison group of 450 patients with established CKD from kidney clinics in the same communities. They compare mortality rates among people who have CKD detected by the new screening test to that of patients who have CKD identified from the kidney clinics. The results are presented below.

	Number of people	Median follow-up (years)	Adjusted mortality rate per 100 person-years ^a
CKD detected by screening test	450	9.1	3.0
CKD from kidney disease clinics	450	5.0	5.5

^aAdjusted for age, race, sex, body mass index, and diabetes

The researchers conclude that the new screening test reduces death from CKD and should be used to screen asymptomatic people for this condition.

90. Which type of bias is most likely responsible for the observed difference in survival?
- A. Lead time bias
 - B. Length bias sampling
 - C. Over diagnosis bias
 - D. Referral bias
 - E. Age bias

The researchers next compare different causes of kidney diseases between the 450 people who have CKD detected by the new screening test and the 450 patients who have CKD identified from the kidney clinics. The results are presented below.

Cause of kidney disease	CKD detected by screening (N = 450)	CKD from kidney disease clinics (N = 450)
Polycystic kidney disease	5%	1%
IgA nephropathy	15%	5%
Hypertensive nephropathy	50%	45%
Diabetic nephropathy	30%	49%

Polycystic kidney disease and IgA nephropathy are typically slowly progressive kidney diseases, whereas diabetic nephropathy is a more rapidly progressive disease.

91. In comparing mortality rates between screened and unscreened individuals, which type of bias is suggested by the above data?
- A. Lead time bias
 - B. Length bias sampling
 - C. Over diagnosis bias
 - D. Referral bias
 - E. Age bias

The researchers next develop a point-of-care urine screening test for CKD. A urine sample is placed on a test strip that changes color along a spectrum from yellow (negative test) to red (positive test). An orange color indicates intermediate results.

The researchers administer the new test to 5000 apparently healthy people. They personally review each of the test strips and enter the results into a database. A total of 670 individuals are found to test positive and are then followed prospectively. Surprisingly, the mortality rate in this group is found to be only 1.5 per 100 person-years, which is lower than that of the group with kidney disease detected by the initial laboratory-based screening test (problem 90).

92. Which type of bias is most likely responsible for the greater survival observed among people who have CKD detected by the new point-of-care test versus that of patients who have CKD identified by the laboratory screening test?
- A. Lead time bias
 - B. Length bias sampling
 - C. Over diagnosis bias
 - D. Referral bias
 - E. Age bias

Chapter 12

Summary Measures in Statistics

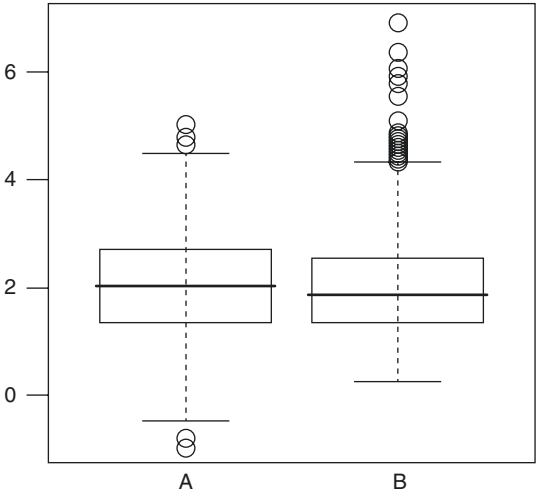


Summary measures provide compact descriptions of one or more study variables. Summary measures include statistical properties, such as the mean and median of a distribution, and graphical presentations, such as histograms and box plots. For normally distributed data, 95% of the observations reside within two standard deviations of the mean value. The joint distribution of two continuous variables can be described graphically using scatter plots, or quantified using correlation, which indicates the tendency of larger values of one variable to match up with larger values of a second variable.

93. A study of plasma thyroid-stimulating hormone (TSH) levels among 10,000 people includes 15 individuals who have extremely high TSH levels. Which of the following summary measures is most likely to be affected by these extreme values?
- A. Mode
 - B. Median
 - C. Mean
 - D. Geometric mean
94. In this same study, plasma TSH levels are roughly normally distributed with a mean value of 3.0 mIU/L and a standard deviation of 0.5 mIU/L. Approximately how many of the 10,000 people in the study are expected to have an TSH level >4.0 mIU/L?
- A. 100
 - B. 250
 - C. 500
 - D. 1000
 - E. 2500

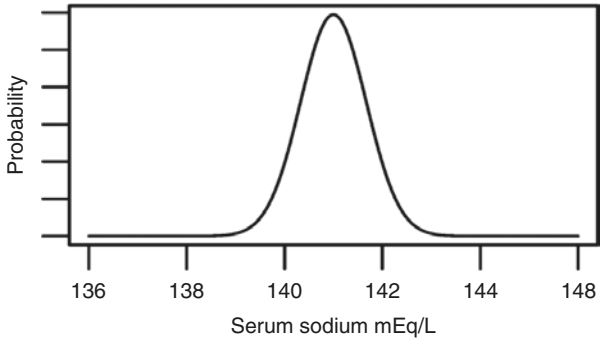
95. The figure below depicts the distributions of two characteristics in a study of 100 people. Which distribution is most likely to have a greater mean value?

- A. Distribution A
- B. Distribution B



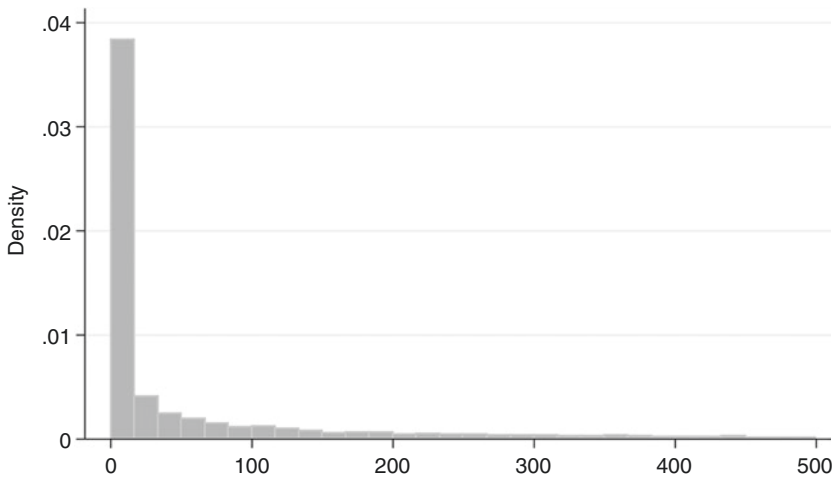
96. The histogram below depicts serum sodium concentrations in an outpatient population. The mean value is 141 mEq/L and the lowest 2.5th percentile of the distribution is 138 mEq/L. What is the approximate standard deviation of this distribution?

- A. 0.5 mEq/L
- B. 0.8 mEq/L
- C. 1.0 mEq/L
- D. 1.5 mEq/L



97. The histogram below presents coronary artery calcium scores, a measure of coronary atherosclerosis, among participants in a large research study. Which one of the following is true?

- A. The mean and median coronary calcium scores are likely to be similar.
- B. A typical person in the study has a high coronary calcium score.
- C. Approximately 95% of participants will have coronary calcium scores within two standard deviations of the mean score.
- D. The mode of coronary calcium scores in the study is 0.
- E. The coronary calcium scores follow a roughly normal distribution.



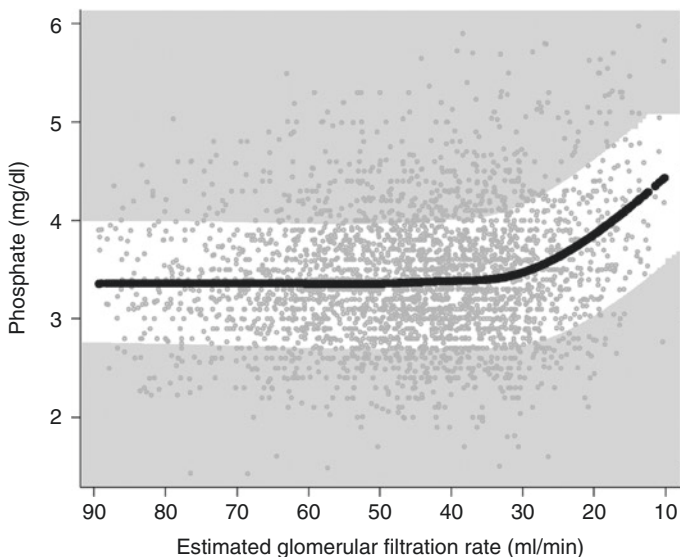
98. Which of the following is an example of a binary variable?

- A. Number of cups of coffee consumed per day
- B. Blood pressure
- C. Age
- D. Cancer stage
- E. Mortality

99. A study evaluating the association between age and usual gait speed reports a correlation coefficient of -0.15 between these characteristics. Which of the following interpretations best describes this result?

- A. Older age tracks with higher gait speeds most of the time.
- B. Older age tends to track with higher gait speeds, but there is considerable variability in this association.
- C. Older age tracks with lower gait speeds most of the time.
- D. Older age tends to track with lower gait speeds, but there is considerable variability in this association.
- E. Age and gait speed are unrelated.

A study evaluates the association between serum phosphate levels and kidney function, estimated by the glomerular filtration rate (GFR). The results are shown below.



The black line in the figure represents mean serum phosphate levels relative to GFR. The gray-shaded areas contain serum phosphate levels in the highest and lowest quintiles relative to GFR, and the white area contains phosphate levels within the three middle quintiles.

100. Which of the following is true?

- A. The 80th percentile of serum phosphate levels is higher for an estimated GFR of 20 ml/min compared with an estimated GFR of 60 ml/min.
- B. Among participants with an estimated GFR of 50 ml/min, about 80% will have a serum phosphate level between 2.8 and 4.0 mg/dl.
- C. The 20th percentile of serum phosphate levels is similar across the measured range of estimated GFR.
- D. The median serum phosphate level is higher for an estimated GFR of 60 ml/min compared with an estimated GFR of 80 ml/min.
- E. The mean serum phosphate level is higher for an estimated GFR of 60 ml/min compared with an estimated GFR of 80 ml/min.

Chapter 13

Statistical Inference



The results obtained in a particular study may or may not reflect those of the larger underlying population. Statistical inference is a mathematical process used to relate findings obtained from a sample (study) to those in the population. Two characteristics that influence how closely sample results are likely to reflect those in the population are sample size and variance. A larger sample size and a smaller variance increase the likelihood that the results obtained in a given study will be indicative of those in the underlying population. *P*-values and 95% confidence intervals are common measures of statistical inference.

A clinical trial compares two antibiotics for the treatment of urinary tract infection caused by *Escherichia coli* (*E. coli*), a common bacterium. The trial recruits 400 women who have acute urinary tract symptoms and a urine culture demonstrating >100,000 *E. coli* colonies per ml. Study personnel randomly assign participants to one of the two study antibiotics and then contact participants after 3 days to encourage completion of these treatments. The primary outcome is the *E. coli* colony count after 7 days of treatment. The results are presented below.

Treatment	Number of participants	Mean end-of-study <i>E. coli</i> colony count (thousands per ml; 95% confidence interval)
Antibiotic one	200	12.7 (8.9, 16.5)
Antibiotic two	200	18.2 (14.4, 22.0)

Difference in mean values = 5.5 thousand colonies per ml
P-value for comparison = 0.04
95% confidence interval (0.2 thousand colonies per ml, 10.8 thousand colonies per ml)

101. Which of the following is true regarding the p -value reported for this study?
- A. If the impact of the two antibiotics on end-of-study colony counts were equal in this study, then the chance of observing a 5.5 thousand colony count difference or an even greater difference in the population is 4%.
 - B. If the impact of the two antibiotics on end-of-study colony counts were different in this study, then the chance of observing the same end-of-study colony counts in the population is 4%.
 - C. If the impact of the two antibiotics on end-of-study colony counts were equal in the population, then the chance of observing a 5.5 thousand colony count difference or an even greater difference in this study is 4%.
 - D. If the impact of the two antibiotics on end-of-study colony counts were different in the population, then the chance of observing similar end-of-study colony counts in this study is 4%.
102. Other than the data shown in the table, what additional information was used to calculate the p -value of 0.04?
- A. Information regarding adherence with the antibiotic treatments
 - B. The proportion of women who were completely cured of *E. coli* in each group
 - C. The mean difference in end-of-study colony counts among all similar women in the underlying population
 - D. The standard deviation for the mean end-of-study colony counts in each group

Consider the following interpretation of the 95% confidence interval reported for this study:

“If the study were repeated an infinite number of times, and a 95% confidence interval constructed for each experimental result, then 95% of these intervals would contain the true difference in the mean end-of-study colony counts, comparing the two antibiotics in the underlying population.”

103. Is this interpretation correct?
- A. Yes
 - B. No
 - C. Cannot determine from the information provided
104. Which of the following represents a valid, though less formal, interpretation of the 95% confidence interval?
- A. There is 95% confidence that the mean end-of-study colony counts differ between the two antibiotics in this study.
 - B. There is 95% confidence that the mean end-of-study colony counts differ by 5.5 thousand colonies per ml between the two antibiotics in this study.

- C. There is 95% confidence that the mean end-of-study colony counts differ by as little as 0.2 or as much as 10.8 thousand colonies per ml in this study.
 - D. There is 95% confidence that the mean end-of-study colony counts differ by as little as 0.2 or as much as 10.8 thousand colonies per ml in the underlying population.
105. Which of the following would result in a narrower confidence interval for the difference in mean end-of-study colony counts comparing the two antibiotics?
- A. Evaluating fewer trial participants
 - B. Reducing the variability in the end-of-study colony counts
 - C. Constructing a 99% confidence interval
 - D. Repeating the study in men

The researchers conduct a second study to assess the real-world impact of antibiotic one. They identify 500 women who have acute urinary tract symptoms from a network of primary care clinics and send an electronic prescription for antibiotic one to their pharmacy. After 5 days of treatment, the mean end-of-study colony count is found to be 19.3 thousand per ml (95% confidence interval (16.9, 21.7)).

106. Which of the following is the most likely conceptual explanation for the relatively weaker impact of antibiotic one on end-of-study colony counts in the second study?
- A. Statistical inference
 - B. Lack of a comparison group
 - C. External validity
 - D. Larger sample size
 - E. Use of a two-sided p -value

Chapter 14

Hypothesis Tests in Practice



Hypothesis tests, such as the t-test, chi-square test, and ANOVA test, yield p -values that represent the probability of observing a particular study result, or a more extreme result, if a pre-specified null hypothesis about the population were true. A p -value threshold, such as 0.05, is typically used to declare statistical significance. Consequently, a hypothesis test may declare a result to be significant when in fact there is no actual difference in the population (type I error) or declare a result to be nonsignificant when in fact there is an actual difference in the population (type II error). Study power, which is the probability of not making a type II error, is influenced by sample size, effect size, variation, and the threshold value for declaring significance.

107. For the first trial comparing the two antibiotics for urinary tract infection (problem 101), which statistical test would be most appropriate for comparing the end-of-study colony counts?
- A. T-test
 - B. ANOVA test
 - C. Chi-square test
 - D. Test for interaction

An exploratory analysis of the first trial evaluates a new outcome: “recovery from urinary tract infection,” which is defined by an end-of-study colony count of <5000 colonies per ml. Results are shown below.

Treatment	Number of participants	Number recovered from infection
Antibiotic one	200	70
Antibiotic two	200	57

P -value for comparison = 0.16

108. Which statistical test would be most appropriate for comparing this outcome between the two antibiotic groups?
- A. T-test
 - B. ANOVA test
 - C. Chi-square test
 - D. Test for interaction

A new test is developed to determine the presence of delirium among patients admitted to the intensive care unit (ICU). Researchers administer the test to 40 consecutive patients at the time of ICU admission and then follow these patients over time to determine their survival status. They find that the presence of delirium at ICU admission is associated with a 15% greater relative risk of death, compared with the absence of delirium; p -value = 0.55.

109. What is the most likely conceptual reason for the statistically nonsignificant result?
- A. Type I error
 - B. Inadequate study power
 - C. Inappropriate statistical significance level
 - D. Multiple comparisons
 - E. Large effect size

Given no significant association of delirium with mortality, the researchers next decide to examine the cross-sectional association of delirium at the time of ICU admission with white matter disease in the hippocampus, determined by magnetic resonance imaging (MRI). This study finds no significant association between delirium and white matter disease; p -value = 0.48.

110. Assume for a moment that the presence of ICU delirium *is* truly associated with hippocampal white matter disease in the population of similar ICU patients. Which of the following would increase the chance of detecting a statistically significant association between delirium and this outcome?
- A. Repeating the study in another group of 40 ICU patients
 - B. Testing the association between delirium and white matter disease in a new group of 20 consecutive patients and then replicating this result in the next 20 patients
 - C. Repeating the study using a more stringent p -value threshold for declaring significance
 - D. Repeating the study after applying strict procedures to standardize the MRI procedure to ensure consistent results from scan to scan

Given no significant association of delirium with white matter disease, the researchers next decide to examine the association of delirium with the length of hospital stay. The study finds that the presence of delirium is associated with a 2-day longer hospital stay, on average, compared with the absence of delirium; p -value = 0.05. Excited by this finding, the researchers attempt to replicate their

findings in the next ten consecutive patients who are admitted to the ICU. The replication study also finds an association of delirium with a 2-day longer hospital stay, but the p -value for this association is 0.30.

111. What is the most likely reason for the statistically nonsignificant result in the replication study?
- A. Type I error
 - B. Inadequate study power
 - C. Inappropriate statistical significance level
 - D. Multiple comparisons
 - E. Large effect size

Chapter 15

Linear Regression



Regression is a mathematical tool for quantifying the association between two or more variables. A fitted univariate linear regression model describes the average difference in one characteristic per unit difference in a second characteristic. Multiple regression is a widely used method for adjusting for several confounding characteristics simultaneously. The fitted multiple regression model can be used to predict the value of the outcome variable for any combination of predictor variables. The individual coefficients from a fitted multiple regression model represent the independent association between the predictor variable and the outcome variable, holding all other variables in the model constant.

An observational study evaluates the association between air pollution levels and lung function among healthy adults aged 30–70 years old. The researchers use neighborhood monitors to determine the amount of exposure to particulate matter <10 μm diameter (PM_{10}). The outcome of the study is the forced expiratory volume in 1 second (FEV_1), which describes the volume of air expelled during the first second of exhalation using spirometry. Results are shown below.

	Unadjusted difference in FEV_1 (ml)	Adjusted difference ^a in FEV_1 (ml)
Age (per year older)	−30	−28
Male	+800	+820
Current smoking	−300	−280
Height (per inch taller)	+370	+100
PM_{10} exposure (per 10 $\mu\text{g}/\text{m}^3$ greater)	−160	−90

^aAdjusted for all of the characteristics in the table

112. For the linear regression equation describing the unadjusted association between PM_{10} exposure and FEV_1 , which value of the beta coefficient for PM_{10} yields the lowest sum of squared residuals?
- A. -30
 - B. +100
 - C. -90
 - D. -160
 - E. -320
113. For the linear regression equation describing the unadjusted association between age and FEV_1 , the value of β_0 (the intercept) is 4400 ml. A 40-year-old person in the study has a measured FEV_1 value of 3700 ml. Based on the unadjusted linear regression model, what is the residual for this person?
- A. -28 ml
 - B. -30 ml
 - C. +30 ml
 - D. +500 ml
 - E. Cannot determine from the information that is provided
114. The average FEV_1 value in the study population is 3 l. Which of the following observations is most likely to influence the slope of the unadjusted linear regression model relating age with FEV_1 ?
- A. A person who is 30 years old and has a measured FEV_1 value of 3 l.
 - B. A person who is 50 years old and has a measured FEV_1 value of 6 l.
 - C. A person who is 70 years old and has a measured FEV_1 value of 6 l.
 - D. A person who is 50 years old and has a measured FEV_1 value of 2 l.
115. Based on the unadjusted regression model relating age with FEV_1 , the predicted FEV_1 value for a newborn (0 years old) is 4400 ml. What is the most likely reason that this predicted value is invalid?
- A. Extrapolation beyond the observed data
 - B. Outlying data points
 - C. Influential data points
 - D. Lack of adjustment for height
 - E. Lack of adjustment for sex
116. Which of the following is true?
- A. After adjustment for age, sex, current smoking, and height, each 1 $\mu g/m^3$ greater PM_{10} exposure is associated with a mean 9-ml lower FEV_1 .
 - B. After adjustment for age, sex, current smoking, and height, each 1 $\mu g/m^3$ greater PM_{10} exposure is associated with a mean 9-ml greater FEV_1 .
 - C. After adjustment for age, sex, current smoking, and height, each 1 $\mu g/m^3$ greater PM_{10} exposure is associated with a mean 9% lower FEV_1 .
 - D. The adjusted analyses demonstrate that greater amounts of PM_{10} exposure cause a reduction in FEV_1 .

117. The p -value for the adjusted association between PM_{10} exposure and FEV_1 is 0.02. Which of the following best describes the interpretation of this p -value?
- A. The probability that PM_{10} exposure is associated with an adjusted difference in FEV_1 in the underlying population is 98%.
 - B. PM_{10} exposure is likely to be associated with a 90 ml/min adjusted difference in FEV_1 in the underlying population.
 - C. Given no adjusted association between PM_{10} exposure and FEV_1 in the underlying population, the chance of observing an adjusted 90 ml/min difference in this study or an even more extreme adjusted association is 2%.
 - D. There is a 98% chance that PM_{10} exposure is associated with a 90 ml/min adjusted difference in FEV_1 in the underlying population.
118. Based on the results in the table, which of the following is true?
- A. Sex, current smoking, height, PM_{10} exposure, or a combination of these characteristics is likely confounding the unadjusted association between age and FEV_1 .
 - B. Age, current smoking, height, PM_{10} exposure, or a combination of these characteristics is likely confounding the unadjusted association between sex and FEV_1 .
 - C. Age, sex, current smoking, height, or a combination of these characteristics is likely confounding the unadjusted association between PM_{10} exposure and FEV_1 .
 - D. The possibility of confounding should be investigated by comparing the p -values for the unadjusted and adjusted associations of PM_{10} exposure with FEV_1 .
 - E. The possibility of confounding should be investigated by comparing the p -values for the unadjusted and adjusted associations for each characteristic in the model.

Chapter 16

Log-Link and Logistic Regression



Regression with log-link or Poisson regression is a model that can be used to study the relative change in an outcome variable. In the log-link regression model, the antilog of each coefficient describes the relative difference in the outcome variable associated with each one-unit difference in the predictor variable. For binary outcomes, the logistic regression model relates the log odds of the outcome with one or more predictor variables. The antilog of each coefficient in a logistic regression model describes the odds ratio of the outcome variable associated with a one-unit difference in the predictor variable.

In a secondary analysis of the air pollution and lung function study, the researchers evaluate the association of PM₁₀ exposure with the proportionate difference in FEV₁. Results are shown below.

	Unadjusted difference in FEV ₁ (%)	Adjusted difference ^a in FEV ₁ (%)
PM ₁₀ concentration (per 10 ug/m ³ greater)	−3.8	−2.4

^aAdjusted for age, sex, current smoking, and height

119. Which regression equation was used to estimate the adjusted relative difference in FEV₁?
- A. $\text{Log (odds FEV}_1) = \beta_0 + \beta_1 \times (\text{age}) + \beta_2 \times (\text{sex}) + \beta_3 \times (\text{smoking}) + \beta_4 \times (\text{PM}_{10})$
 - B. $\text{Log (mean FEV}_1) = \beta_0 + \beta_1 \times (\text{age}) + \beta_2 \times (\text{sex}) + \beta_3 \times (\text{smoking}) + \beta_4 \times (\text{PM}_{10})$
 - C. $\text{Mean (FEV}_1) = \beta_0 + \beta_1 \times (\text{age}) + \beta_2 \times (\text{sex}) + \beta_3 \times (\text{smoking}) + \beta_4 \times (\text{PM}_{10})$

- D. $\text{Odds FEV}_1 = \beta_0 + \beta_1 \times (\text{age}) + \beta_2 \times (\text{sex}) + \beta_3 \times (\text{smoking}) + \beta_4 \times (\text{PM}_{10})$
- E. $\text{Log (relative risk FEV}_1) = \beta_0 + \beta_1 \times (\text{age}) + \beta_2 \times (\text{sex}) + \beta_3 \times (\text{smoking}) + \beta_4 \times (\text{PM}_{10})$
120. Consider two people in the study who are the same age, sex, and height and have the same smoking status. Person one has a PM_{10} exposure of 20 ug/m^3 . Based on the adjusted model, this person has a predicted FEV_1 value of 3900 ml. Person two has a PM_{10} exposure of 30 ug/m^3 . What is the predicted FEV_1 value for person two?
- A. 3898 ml
- B. 3809 ml
- C. 3745 ml
- D. 3660 ml
- E. 3200 ml

The researchers conducting the air pollution study next decide to assess the association of PM_{10} exposure with a clinically relevant outcome: asthma. Results are presented below.

	Odds ratio of asthma (95% confidence interval)	
	Unadjusted	Adjusted ^a
PM_{10} exposure (per 10 ug/m^3 greater)	1.8 (1.5, 2.2)	1.4 (1.1, 1.8)

^aAdjusted for age, sex, current smoking, and height

121. Which of the following represents a valid interpretation of the adjusted association?
- A. The adjusted odds of asthma is 1.4, holding age, sex, current smoking, and height constant.
- B. Exposure to any PM_{10} is associated with a 40% greater odds of asthma, holding age, sex, current smoking, and height constant.
- C. A PM_{10} exposure of 10 ug/m^3 is associated with a 40% odds of asthma, holding age, sex, current smoking, and height constant.
- D. Each 10 ug/m^3 greater PM_{10} exposure is associated with a 40% higher odds of asthma, holding age, sex, current smoking, and height constant.
122. Which of the following represents a valid interpretation of the confidence interval for the adjusted association between PM_{10} exposure and asthma?
- A. If the study was repeated an infinite number of times, we can be 95% confident that the adjusted odds ratio of asthma associated with each 10 ug/m^3 greater PM_{10} exposure would be between 1.1 and 1.8 each time.
- B. If the study was repeated an infinite number of times, and a 95% confidence interval placed around each adjusted odds ratio, then 95% of these adjusted odds ratios would be between 1.1 and 1.8.

- C. We can be 95% confident that the adjusted odds ratio of asthma associated with each 10 ug/m^3 greater PM_{10} exposure is as low as 1.1 or as high as 1.8 in the underlying population of people similar to those in this study.
- D. We can be 95% confident that the adjusted odds ratio of asthma associated with each 10 ug/m^3 greater PM_{10} exposure is as low as 1.1 or as high as 1.8 in this study.

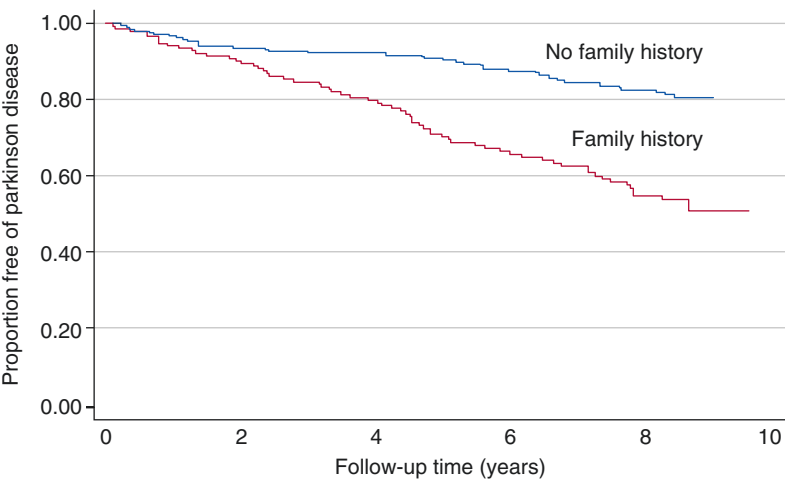
Chapter 17

Survival Analysis



Survival analysis is used to describe the continuous probability of disease-free survival over follow-up. The survivor function, $S(t)$, is a function fit to the study data that returns the cumulative probability of being free of the outcome at a particular time. The Kaplan-Meier method is used to estimate $S(t)$ in the presence of censoring, which is defined as leaving a study for any reason before incurring the outcome of interest. Kaplan-Meier plots are useful for estimating time-specific survival among one or more study groups. The Cox proportional hazards model yields a measure of risk called the *hazard ratio*, which represents an average relative risk over a specified period of follow-up time.

A prospective cohort study examines the association of a family history of Parkinson disease with the development of this disease over time. The investigators recruit participants who are initially free of known Parkinson disease at the start of the study and then follow them until they either develop Parkinson disease, drop out of the study, die, or the study ends, whichever comes first. The Kaplan-Meier plot below describes the development of Parkinson disease over the course of the study. The lower line represents participants who have a family history of Parkinson disease, and the upper line represents participants who do not have a family history of the disease.



123. Which of the following is true?
- A. Among people who have a family history of Parkinson disease, approximately 80% developed the disease after 4 years of follow-up.
 - B. The survival curves account for potential differences in age between people with and without a family history of Parkinson disease.
 - C. The median time to Parkinson disease was approximately 10 years greater among people with a family history compared to those without a family history.
 - D. Participants were censored in this study when they either developed Parkinson disease, dropped out, died, or the study ended.
 - E. After 8 years of follow-up, the difference in Parkinson disease-free survival was about 27%, comparing people with a family history to those without a family history.

The following table presents hazard ratios for developing Parkinson disease during the study.

Family history of Parkinson disease	Unadjusted hazard ratio (95% confidence interval)	Adjusted hazard ratio ^a (95% confidence interval)
No	Reference	Reference
Yes	2.0 (1.5, 2.5)	2.2 (1.6, 2.8)

^aAdjusted for race, sex, and smoking

124. Which of the following is true?

- A. The unadjusted relative risk of Parkinson disease is two times higher comparing people with a family history of the disease to those without a family history.
- B. Unadjusted survival with Parkinson disease is 2 years longer comparing people with a family history of the disease to those without a family history.
- C. Adjusted survival with Parkinson disease is 2.2 times higher comparing people with a family history of the disease to those without a family history holding race, sex, and smoking status constant.
- D. Death due to Parkinson disease is 2.2 times higher comparing people with a family history of the disease to those without a family history holding race, sex, and smoking status constant.
- E. The hazard ratios fail to account for potential differences in follow-up time between participants with and without a family history of Parkinson disease.

A new synthetic endovascular graft is developed for use in high-risk vascular surgeries. A large randomized trial compares the new graft to standard endovascular grafts among 2000 patients who are assessed to be at high postoperative risk. The researchers randomly assign participants to receive either the new endovascular graft or a standard graft. The outcome of the study is all-cause mortality and drop out is minimal. Results over 1 year of follow-up are presented below.

Postoperative period	Number of deaths	Hazard ratio comparing new synthetic graft to standard endovascular graft
Day 0–30	160	3.0
Day 31–60	240	0.5
Day 61–90	180	0.5
Day 91–365	120	0.5

125. Approximately when would the survival curves be expected to cross?

- A. 15 days postoperatively
- B. 30 days postoperatively
- C. 60 days postoperatively
- D. 90 days postoperatively
- E. 180 days postoperatively

Chapter 18

Practice Problem Workbook Solutions



The following table summarizes the rates and proportions calculated from the participant-level data. Recall that incidence is defined as the number of new cases of disease that occur among people who do not have that disease at the beginning of a study. Consequently, the follow-up time column is calculated by adding the individual follow-up times among only people *without* diabetes at the start of the study.

Beta-carotene use	Number of people	Diabetes at baseline	New diabetes cases during follow-up ^a	Follow-up time (months) ^a
Yes	20	4	5	706
No	20	2	5	636
Overall	40	6	10	1342

^aIncidence data calculated among participants without diabetes at baseline

1. Prevalence of diabetes overall = $\frac{\text{Number with diabetes at baseline}}{\text{number in population}} \times 100\%$
 $= 6 / 40 \times 100\% = 15\%$
2. Incidence proportion overall = $\frac{\text{New cases of diabetes}}{\text{number initially free of diabetes}} \times 100\%$
 $= 10 / 34 \times 100\% = 29.4\%$
3.
$$\text{Incidence proportion (beta - carotene)} = \frac{\text{New cases of diabetes among beta carotene users}}{\text{Number of beta carotene users initially free of diabetes}} \times 100\%$$

$$= 5 / 16 \times 100\% = 31.3\%$$

$$\begin{aligned}
 4. \quad \text{Incidence rate (beta-carotene)} &= \frac{\text{New cases of diabetes among beta carotene users}}{\text{Person-time at risk among beta carotene users}} \\
 &= 5 / 706 = 0.0071 \text{ cases per person-month. Round to 1000 :} \\
 &= 7.1 \text{ cases per 1000 person-months}
 \end{aligned}$$

$$\begin{aligned}
 5. \quad \text{Incidence rate (no beta-carotene)} &= \frac{\text{New cases of diabetes among non-beta carotene users}}{\text{Person-time at risk among non-beta carotene users}} \\
 &= 5 / 636 = 0.0079 \text{ cases per person-month.} \\
 &\quad \text{Round to 1000 :} \\
 &= 7.9 \text{ cases per 1000 person-months}
 \end{aligned}$$

6. A. Prevalence = (incidence) $\times$ (duration). In this example, the incidence of disease remains constant, but the duration of the disease increases; therefore, prevalence will increase. Prevalence provides a snapshot of the amount of disease that is present at a given time. The widespread use of antiviral therapies for HIV has dramatically increased the lifespan of people who contract this disease. Consequently, the proportion of people living with HIV disease in a given snapshot of time has increased.

$$7. \quad C. \text{ Prevalence (men)} = \text{Incidence (men)} \times \text{Duration (men)}$$

$$\text{Prevalence (women)} = \text{Incidence (women)} \times \text{Duration (women)}$$

The incidence of the disease is five times greater in men compared to women. For the prevalences to be equal, the duration of the disease must be five times *smaller* in men. This could be due to men dying from the disease more rapidly, for example, the disease may be more aggressive in men, or could be due to men recovering more quickly, for example, men may receive more intensive treatment for the disease.

8. C. This study determines the *prevalence* of irritable bowel syndrome (IBS) symptoms. Specifically, the study determines the amount of IBS symptoms that are present at the time the questionnaire is administered. A study to assess the incidence of IBS symptoms would require starting with a population that is initially free of such symptoms and then determining the development of new symptoms over time.

The prevalence of IBS symptoms among individuals who have anxiety disorder in this study is:

$$\text{Prevalence} = \frac{\text{Number with anxiety disorder and IBS symptoms}}{\text{Number with anxiety disorder}} = \frac{4000}{10,000} \times 100\% = 40\%$$

9. A. Prevalence measures are useful for raising awareness of a disease and guiding allocation of health resources. The study data indicate that IBS symptoms are common among people who have anxiety disorder, suggesting a high-risk population for which provision of educational materials regarding IBS may be useful. The data do not clarify whether the anxiety disorder preceded the IBS symptoms or vice versa, limiting causal inference and precluding a change in clinical practice.
10. C. Incidence rates based on person-hours would provide the fairest comparison of the incidence of carpal tunnel syndrome between the two companies. It is possible that the competitor company hires more part-time employees, who work fewer hours, resulting in a lower incidence proportion of carpal tunnel syndrome.
11. The study population is infants (including those with and without hyperbilirubinemia).
The exposure is hyperbilirubinemia.
The outcome is language delay at 3 years.
12. The study population is students who attend one of the five metropolitan high schools.
The exposure is the amount of daily caloric intake.
The outcome is creatine supplement use.
13. The study population is general surgeons.
The exposure is the number of previous laparoscopic cholecystectomy procedures.
The outcome is bile leak.
14. D. General categories of exclusion criteria include prevalent disease, major risk factors for the disease, inability to obtain valid study data, and safety (in a clinical trial). Excluding patients who have a previous history of the study outcome, bile leak, or major risk factors for this outcome, would be reasonable. Moreover, excluding surgeons who lack available data to determine their exposure status would also be reasonable. However, there is no clear rationale for excluding surgeons who performed low numbers of cholecystectomy procedures; such surgeons provide important information about the study hypothesis.
15. D. Internal validity addresses whether a study correctly answers the question of interest within the specified population and environment. Some characteristics that promote the internal validity of a study include a population that is free of other risk factors for the disease, to focus on a specific risk factor of interest, accurate measurements of the study data, randomization, and an appropriate analytic plan. Medical chart review would provide an accurate determination of the study outcome, promoting internal validity.
16. B. External validity addresses whether the results of a study can be generalized to broader groups of people and environments. Evaluating the association of surgical experience with bile leak within a single healthcare organization will generate results that are applicable to the practice patterns, training procedures,

and patient and surgeon characteristics of that organization. Conducting the study across multiple healthcare organizations would yield findings that are more broadly applicable.

17. B – Biologic plausibility. Case series present the experiences of a group of people who have a specific disease or condition. Case series are important for raising awareness of a disease and carefully describing its salient features; however, this design is typically inadequate for inferring causal relationships. Ideally, the association between complement factor H levels and macular degeneration would be assessed by comparing the *incidences* of this disease among people who have low versus high levels of complement factor H. The case series data do not include the necessary information to calculate such incidences. Moreover, the case series data do not distinguish whether low complement factor H levels preceded the development of macular degeneration or whether low complement factor H levels occurred later, as a consequence of this condition. A possible causal role of low complement factor H levels in the development of macular degeneration is supported by biologic plausibility, based on knowledge from other studies.
18. C – Biologic plausibility and temporal relationship. This case series describes a group of people who developed cardiac arrhythmia while receiving antipsychotic medications. No information is provided regarding the total number of antipsychotic medication users from whom these arrhythmia cases developed, precluding determination of the association with disease. These study data *do* demonstrate a temporal relationship between antipsychotic medication use and arrhythmia because the arrhythmias occurred after the initiation of these medications. There is biologic plausibility for a causal effect of antipsychotic medications on torsades de pointes based on their ability to prolong ventricular repolarization.
19. E. A denominator is necessary to calculate the *rates* of defective tires at each plant. Observing a greater raw number of defective tires originating from the Cincinnati plant may simply reflect a (much) greater overall output of this plant. The other factors listed, such as characteristics of the workers and characteristics of the specific plants, may be important for subsequently understanding the reasons *why* a particular plant might produce more defective tires. However, first things first – rates are needed to determine *whether* the plants actually differ in their production of defective tires.
20. D. The case series data describe opportunistic infections among only users of the new medication. Such infections are not infrequent among transplant recipients who typically receive many immunosuppressant medications. Rates of opportunistic infection among users of the new medication should be compared with rates of opportunistic infection among transplant recipients who do not receive this medication.
21. B. Table 1 of the paper describes findings from the cross-sectional component of the study. Cross-sectional ascertainment of the exposure and outcome permits calculation of prevalences, which cannot clarify temporal relationships. For example, an association between having children and serious thoughts of

dropping out of medical school does not specify whether the child was born before or after the student developed thoughts of dropping out – both scenarios are plausible. Based on the data in Table 1:

Prevalence of serious thoughts of dropping out among students who have children = 18.1%

Prevalence of serious thoughts of dropping out among students who do not have children = 10.0%

Relative prevalence = $18.1\%/10.0\% = 1.8$.

22. D. Based on the data in Table 2 of the paper, there are 195 students with burnout among the group with serious thoughts of dropping out and another 856 students with burnout among the group without serious thoughts of dropping out, yielding a total of 1051 students with burnout in the cross-sectional sample. In this same sample, the denominator for prevalence is the total number of students who provided data regarding burnout, which is $195 + 856 + 40 + 1038 = 2129$.

Prevalence of burnout = $1051 / 2129 \times 100\% = 49.4\%$

23. E. Incidence rate is defined as the number of new cases of disease divided by person-time at risk. New cases of disease are typically defined as those that occur among people who did not have the condition at the start of the study. To determine the incidence of serious thoughts of dropping out of medical school, a group of students without such thoughts at the start of the study would be needed. The paper does not provide data regarding how many students developed “new” thoughts of dropping out from 2006 to 2007.
24. D. The association of burnout with serious thoughts of dropping out of medical school is relatively strong, evidenced by an odds ratio = 5.9 (Table 2). There is obvious plausibility for such an association. Table 2 also demonstrates a “dose-varying” relationship; the first three rows of this table indicate that progressively greater burnout scores are associated with a stepwise higher risk of serious thoughts of dropping out. Clearly this is not a randomized trial; there is no practical or ethical way to assign students to feelings of burnout. The cross-sectional study design cannot disentangle the temporal relationship between feelings of burnout and serious thoughts about dropping out of medical school; either one may have preceded the other.
25. A. This cross-sectional study ascertains race and serum 25-hydroxyvitamin D levels at the same time. However, temporal relationships can sometimes be inferred from cross-sectional data if one of the directions of association is impossible or extremely unlikely. In this example, a person’s race clearly precedes their circulating level of vitamin D. The opposite possibility, that circulating vitamin D levels influence a person’s race, is implausible.
26. B. The situation is less clear for the cross-sectional analysis of obesity and vitamin D levels. On one hand, low serum vitamin D levels may precede the development of obesity (and possibly contribute to the pathogenesis of this

condition). On the other hand, obesity can lead to sequestration of vitamin D within fat tissue, potentially reducing circulating vitamin D levels. Cross-sectional ascertainment of serum vitamin D levels and obesity prevents separation of these two possibilities.

27. E. A cohort study that begins with nonobese individuals, measures serum 25-hydroxyvitamin D levels at the start of the study, and then determines new (incident) cases of obesity over time would help clarify the temporal relationship between these characteristics. Nonetheless, an association between low baseline vitamin D levels and incident obesity may not indicate a causal impact of vitamin D deficiency on this outcome, because people who have low versus high vitamin D levels may differ by other characteristics that could also affect the incidence of obesity, such as physical activity levels and the amount of time spent outdoors.
28. A. This is a cohort study, in which people who are free of the outcome (hypoglycemia) at the start of the study are followed over time for the development of new instances of the outcome. The exposure is the type of insulin therapy (long-acting versus short-acting), which is measured, but not assigned, by the researchers. Randomized trials have a similar design but *assign* participants to specific treatments or control procedures. The cohort design improves upon the case series design by including groups of exposed and unexposed people and improves upon the cross-sectional design by determining new instances of the outcome (hyperglycemia) over time among people who do not initially have this condition.
29. C. Incidence rates are preferable to incidence proportions for calculating relative risk. Determination of incidence rates requires adding all of the person-time in each cohort.

$$\begin{aligned}\text{Incidence rate (long-acting insulin)} &= 2 \text{ events} / 9.8 \text{ person-years} \\ &= 20.4 \text{ events per 100 person-years}\end{aligned}$$

$$\begin{aligned}\text{Incidence rate (short-acting insulin)} &= 2 \text{ events} / 8.2 \text{ person-years} \\ &= 24.4 \text{ events per 100 person-years}\end{aligned}$$

$$\text{Relative risk} = 20.4 / 24.4 = 0.84 \text{ (no units)}$$

30. B. Attributable risk describes the additional amount of disease attributable to a particular risk factor among exposed individuals.

$$\begin{aligned}\text{Incidence rate (low pollution)} &= 200 \text{ asthma cases} / 3300 \text{ person-years} \\ &= 6.1 \text{ cases per 100 person-years}\end{aligned}$$

$$\begin{aligned}\text{Incidence rate (high pollution)} &= 100 \text{ asthma cases} / 780 \text{ person-years} \\ &= 12.8 \text{ cases per 100 person-years}\end{aligned}$$

$$\begin{aligned}
 \text{Attributable risk} &= \text{Incidence}(\text{exposed}) - \text{Incidence}(\text{unexposed}) \\
 &= \text{Incidence}(\text{high pollution}) - \text{Incidence}(\text{low pollution}) \\
 &= 6.7 \text{ cases per 100 person - years}
 \end{aligned}$$

There are 6.7 additional asthma cases per 100 person-years among children who are exposed to high levels of air pollution. If high air pollution levels are the sole cause of the difference in asthma rates observed in this study, then these extra cases of asthma are attributable to this environmental exposure.

31. A. Population attributable risk describes the additional amount of disease attributable to a particular risk factor in a population.

$$\begin{aligned}
 \text{Incidence rate}(\text{population}) &= 300 \text{ asthma cases} / 4080 \text{ person - years} \\
 &= 7.4 \text{ cases per 100 person - years}
 \end{aligned}$$

$$\begin{aligned}
 \text{Population attributable risk} &= \text{Incidence}(\text{population}) - \text{Incidence}(\text{unexposed}) \\
 &= \text{Incidence}(\text{population}) - \text{Incidence}(\text{low pollution}) \\
 &= 7.4 \text{ cases per 100 person - years} - 6.1 \text{ cases per} \\
 &\quad 100 \text{ person - years} \\
 &= 1.3 \text{ cases per 100 person - years}
 \end{aligned}$$

There are 1.3 additional asthma cases per 100 person-years among children in this population. Given the assumption that high levels of air pollution are the sole cause of the observed difference in asthma rates, removing this exposure would be expected to reduce the incidence of asthma in the population by 1.3 cases per 100 person-years.

32. D. This is a pharmacoepidemiology study that evaluates cohorts of atypical and conventional antipsychotic medication users. The study population consists exclusively of people who use an antipsychotic medication (either atypical or conventional). The design complements that of randomized trials by facilitating assessment of relatively large numbers of people in a real-world setting. Many people in this study had comorbid illnesses and mental health conditions that may have precluded their participation in a randomized trial (Table 1 of the paper). Moreover, trials typically evaluate medication use under stringent monitoring procedures that may not be representative of those in real-world settings. The primary limitation of this observational study is the inability to clearly separate the causal impact of the antipsychotic medication types on mortality from potential differences in the characteristics of people who use these different medications.
33. C. The study does not include person-time information for calculation of incidence rates. Therefore, incidence proportions must be used to calculate attributable risk. The incidence proportion of all-cause mortality within 180 days is presented in the last row of Table 1.

Incidence proportion (conventional antipsychotic medication use) = 17.9%

Incidence proportion (atypical antipsychotic medication use) = 14.6%

$$\begin{aligned}\text{Attributable risk} &= \text{Incidence}(\text{exposed}) - \text{Incidence}(\text{unexposed}) \\ &= 17.9 - 14.6\% \\ &= 3.3\%\end{aligned}$$

This attributable risk can be interpreted as, “if the use of conventional antipsychotic medications (as compared with atypical agents) were the sole cause of the mortality difference observed in this study, then there are 3.3% additional deaths within 180 days among conventional antipsychotic medication users that are attributable to this medication class.”

34. C – Dose-response and temporal relationship. The observed association between conventional antipsychotic medication use and mortality in this study is *not* particularly strong: the relative risk (hazard ratio) is less than 1.5. The authors admit that biological mechanisms to explain *why* conventional antipsychotic medications may be more dangerous than atypical agents are lacking (discussion section). Some evidence of a dose-response is shown in Table 2, in which higher doses of conventional antipsychotic medications are associated with greater risks of mortality. The cohort study design is observational and therefore does not represent randomized evidence. However, the study does demonstrate that the type of antipsychotic medication use preceded the occurrence of mortality.
35. C. This is a case-control study. The study begins by selecting case individuals who have the disease of interest (multiple sclerosis) and a group of control individuals who do not have this disease. The odds ratio is the primary measure of risk in case-control studies. Calculation of the odds ratio requires creation of a 2×2 table of exposure and disease status shown below.

	Disease (multiple sclerosis)	No disease (no multiple sclerosis)
Exposed (hepatitis B vaccination)	a (12)	b (13)
Unexposed (no hepatitis B vaccination)	c (88)	d (87)
Total	100	100

$$\text{Odds ratio} = (a \times d) / (b \times c) = (12 \times 87) / (13 \times 88) = 0.91$$

It is *not* possible to estimate specific incidences of multiple sclerosis among hepatitis B vaccinated or non-vaccinated individuals from the case-control study data alone, because this design artificially fixes the ratio of cases and controls. Similarly, it is not possible to estimate attributable risk or population attributable risk directly from case-control data. However, it is possible to *estimate the relative risk* of multiple

sclerosis, comparing people who received the hepatitis B vaccine to those who did not receive this vaccine. This estimate is called the odds ratio, which approximates the relative risk if the frequency of the disease is low in the underlying population.

36. A. To increase the chance of obtaining valid associations, case-control studies should ideally select case individuals based on a high degree of certainty for having the disease of interest. The use of magnetic resonance imaging to define multiple sclerosis would increase the specificity of the disease and enhance the internal validity of the study. Selection of healthcare workers as controls would artificially exaggerate the frequency of hepatitis B vaccination relative to the cases; selection of uninsured adults as controls would artificially reduce this frequency. There is no possibility of recall bias in this study because the exposure, hepatitis B vaccination, was determined by review of medical records.
37. C. The data needed to calculate this odds ratio are found in Table 1 of the paper.

	Disease (bladder cancer)	No disease (no bladder cancer)
Exposed (cyclophosphamide, no radiation)	9	20
Unexposed (no cyclophosphamide, radiation)	6	16

Odds ratio = $(a \times d)/(b \times c) = (9 \times 16)/(20 \times 6) = 1.2$

38. E. It is not possible to determine the incidence of the disease (bladder cancer) among exposed or unexposed people using case-control data alone.
39. A. Confirmation of case status (secondary bladder and kidney cancers) is an important desirable feature of case-control studies. The investigators used previously collected medical records to avoid recall bias that could have resulted from interviewing participants, who would have likely been aware of a bladder or kidney cancer diagnosis. Case and control participants were selected from the same cancer registries. Bladder and kidney cancers are relatively uncommon in the population, permitting estimation of relative risks from odds ratios. There is no harm in confirming the original diagnosis of non-Hodgkin lymphoma in this study. However, such confirmation does not bear directly on the validity of associations between chemotherapy and radiation and the development of secondary kidney or bladder cancers.
40. D. Table 1 of the paper presents the numbers of case and control subjects who received cyclophosphamide with and without radiation. Table 3 of the paper presents numbers of case and control subjects who received no radiotherapy and no cyclophosphamide. These data are summarized below.

	Disease (bladder cancer)	No disease (no bladder cancer)
Cyclophosphamide, no radiation	9	20
Cyclophosphamide, radiation	9	10
No radiation, no cyclophosphamide	6	42

Using the no radiation, no cyclophosphamide group as a reference:

Odds ratio of bladder cancer associated with cyclophosphamide
and no radiation = $(9 \times 42) / (20 \times 6) = 3.2$

Odds ratio of bladder cancer associated with cyclophosphamide
plus radiation = $(9 \times 42) / (10 \times 6) = 6.3$

Table 4 of the paper shows that both dose and duration of cyclophosphamide are associated with secondary bladder cancer. Table 2 shows that the dosage of radiation is in fact slightly lower among bladder cancer cases compared to controls. The data in Table 1 can be used to calculate the odds ratio of bladder cancer comparing men to women; men have a 2.3 greater odds of bladder cancer compared to women in this study.

41. E. The study investigators were blinded to the treatment assignments, but not the study participants. There is no practical way to prevent participants from knowing whether they are receiving an exercise program or not. The lack of blinding could disproportionately influence participant's behaviors outside of the trial. For example, people who were assigned to no exercise (and were aware of this assignment) may have been more likely to engage in exercise on their own. The large randomized trial design will create similar distributions of age, sex, fasting blood sugar, family history of diabetes, and other characteristics across the treatment groups, providing reassurance that such characteristics will *not* bias the trial results.
42. E. Internal validity describes the ability of a study to accurately determine the benefits and harms of the treatments within the selected population and environment. Low adherence with the prescribed exercise program (or initiation of exercise in the control group) could reduce the ability of the trial to correctly determine the impact of this exercise itself on the incidence of diabetes. Exclusion for heart failure influences external validity because the trial results do not apply to people with a history of heart failure. Diabetes is a clinical outcome, not a surrogate endpoint. There is no need to employ complex randomization procedures in such a large trial (3400 people); balance in characteristics should be achieved by sample size alone.
43. A. External validity addresses whether the results of a study are likely to be broadly applicable to more general groups of people and realistic environments. Frequently contacting trial participants to encourage adherence with exercise could increase the internal validity of the study, if such procedures improved compliance with the exercise program. On the other hand, real-world participation in exercise programs does not typically include scheduled reminders. The impact of exercise programs on the development of diabetes could be greater in the setting of a trial, in which this treatment is vigorously promoted, compared to more realistic settings.

44. E.

$$\text{Relative risk} = \frac{\text{Incidence}(\text{treated})}{\text{Incidence}(\text{untreated})} = \frac{6.5 \text{ events per } 1000 \text{ person} - \text{years}}{8.7 \text{ events per } 1000 \text{ person} - \text{years}} = 0.75 (\text{no units})$$

45. B. Attributable risk = Incidence(treated) – Incidence(untreated)

$$\begin{aligned} &= 6.5 \text{ events per } 1000 \text{ person} - \text{years} \\ &\quad - 8.7 \text{ events per } 1000 \text{ person} - \text{years} \\ &= -2.2 \text{ events per } 1000 \text{ person} - \text{years} \end{aligned}$$

46. E. Number needed to treat or harm can be calculated using attributable risks derived from incidence rates or proportions. For ease of interpretation, incidence proportions will be used to calculate attributable risks over a median of 5 years of follow-up in the trial.

$$\begin{aligned} \text{Attributable risk} &= \text{Incidence}(\text{treated}) - \text{Incidence}(\text{untreated}) \\ &= 3.2 \text{ events per } 100 \text{ people} - 4.4 \text{ events per } 100 \text{ people} \\ &= -1.2 \text{ events per } 100 \text{ people} \end{aligned}$$

$$\begin{aligned} \text{Number needed to treat} &= 1 / \text{absolute value}(\text{attributable risk}) \\ &= 1 / \text{absolute value}(-1.2 \text{ events per } 100 \text{ people}) \\ &= 83 \text{ people} / \text{event} \end{aligned}$$

Treating 83 people who are similar to those in the trial with the exercise program, instead of no treatment, would be expected to prevent one case of diabetes over a median of 5 years of follow-up.

47. B. In randomized trials, attributable risk describes the additional burden of outcomes among the untreated group. Based on the solution to problem 45, there were 2.2 extra cases of diabetes per 1000 person-years among trial participants who were assigned to no exercise. The large randomized design provides reassurance that the characteristics of participants who were assigned to the exercise intervention were similar to those of participants assigned to no exercise. Therefore, the additional cases of diabetes are likely due to the lack of the exercise intervention itself. Providing this same exercise program to a large number of people who are like those in the trial and not currently exercising would be expected to reduce the incidence rate of diabetes by 2.2 cases per 1000 person-years.

48. C. The large randomized trial design will create similar distributions of characteristics among participants assigned to the exercise program and participants assigned to no treatment *at the start of the trial*. If these groups are analyzed according to the initial randomization assignment (intention-to-treat), then observed differences in diabetes outcomes are likely to reflect the impact of the

exercise program itself. Other analytic methods may disturb the initial balance in characteristics created by randomization and conflate the impact of treatment with the characteristics of people who tend to adhere to the treatment or those who initiate the treatment outside of the trial.

49. B. Despite the promising preliminary experimental data, the subgroup analyses do *not* support a differential impact of the exercise program on diabetes incidence comparing participants who were sedentary versus those who were non-sedentary at the start of the trial. Based on the trial data:

$$\text{Relative risk (sedentary behavior group)} = 8.0 / 10.3 = 0.78$$

$$\text{Relative risk (non – sedentary behavior group)} = 5.6 / 7.8 = 0.72$$

$$\begin{aligned} \text{Attributable risk (sedentary behavior group)} \\ = 8.0 - 10.3 = -2.3 \text{ per 1000 person – years} \end{aligned}$$

$$\begin{aligned} \text{Attributable risk (non – sedentary behavior group)} \\ = 5.6 - 7.8 = -2.2 \text{ per 1000 person – years} \end{aligned}$$

The incidence rates of diabetes are higher among sedentary participants; however, the impact of the exercise intervention, measured on either the relative or absolute risk scale, is similar.

50. C. Locally advanced breast cancer is a serious disease that requires immediate treatment. It would be unethical to assign women who have locally advanced cancer to no treatment at all, a placebo, or a delay in treatment. The control group in a trial targeting locally advanced breast cancer should receive the accepted standard of care for the disease. The addition of a placebo to standard of care treatment will blind the study participants as to whether they are receiving the new medication.
51. C. External validity describes the ability to apply the results of a study to more general groups of people in real-world environments. *The study population and the study environment impact external validity.*

The lack of blinding, use of a subjective outcome, and potential nonadherence with treatment may impact internal validity, which describes the ability of a trial to correctly answer the proposed question of interest.

52. E. Internal validity addresses the likelihood that a study truthfully answers the clinical or scientific question within the selected study population and environment. Large, well-conducted randomized trials of clinically relevant outcomes tend to have high internal validity. However, not all trials are internally valid, and other study designs may also yield internally valid results. This trial addressed the specific question: does denosumab treatment prevent fractures among postmenopausal women with osteoporosis? The answer to this question is yes. Randomization of large numbers of participants resulted in nearly identical characteristics between the denosumab and placebo groups (Table 1 of the paper). The only meaningful difference between these groups at the time of randomization is the use versus non-use of denosumab itself (although some

participants assigned to denosumab stopped this treatment during the trial and some assigned to placebo may have initiated denosumab outside of the trial). Under an intention-to-treat analysis, the difference in fracture incidence is almost certainly due to denosumab treatment. The primary outcome of the trial was fracture, which is a clinical outcome, not a surrogate endpoint. Exclusion for long-term bisphosphonate use narrows the external validity of the trial results to women who do not use bisphosphonates.

53. C. Incidence proportion (treated) = $43 / 3906 \times 100 = 1.10$ events per 100 people

Incidence proportion (untreated) = $26 / 3902 \times 100 = 0.67$ events per 100 people

$$\begin{aligned}\text{Attributable risk} &= \text{Incidence}(\text{treated}) - \text{Incidence}(\text{untreated}) \\ &= 0.67 \text{ events per 100 people} - 1.10 \text{ events per 100 people} \\ &= -0.43 \text{ events per 100 people}\end{aligned}$$

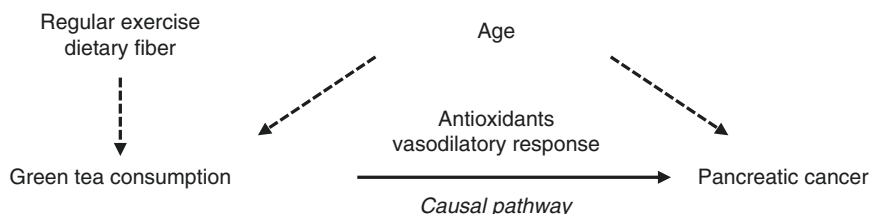
$$\begin{aligned}\text{Number needed to treat} &= 1 / \text{absolute value}(\text{attributable risk}) \\ &= 1 / (0.43 \text{ events per 100 people}) = 232.6\end{aligned}$$

Approximately 233 postmenopausal women with osteoporosis who are similar to those in the trial would need to be treated with denosumab to prevent one hip fracture within 36 months.

54. C. Subtle risk signals detected in trials are often difficult to interpret due to small numbers of events and multiple comparisons among the many different adverse outcomes. Nonetheless, there were 12 times as many occurrences of cellulitis among women who were assigned to denosumab compared with those assigned to placebo. The randomized trial design increases the degree of certainty that these extra cellulitis cases were caused by denosumab treatment itself; however, chance remains a possibility. There is some plausibility for an adverse impact of denosumab on cellulitis as this medication affects innate immunity via a major cell signalling pathway (RANK ligand system). If the observed difference in cellulitis outcomes was in fact caused by denosumab, rather than chance, then widespread use of denosumab could result in many excess cases of cellulitis in the population. The most practical approach to the trial findings is post-approval surveillance, in which adverse effects continue to be monitored after a drug is approved. The safety signals observed in the trial would ideally be addressed by evaluating denosumab treatment among larger, more inclusive study populations in more real-world environments.
55. A. The intention-to-treat approach is typically the preferred method for analyzing randomized trial data because this method preserves the initial balance in characteristics created at randomization. Other analytic methods that permit switching or stopping treatment once the trial is underway are no longer protected from potential imbalances in participants characteristics and are therefore susceptible to confounding by factors that may prompt switching or stopping treatment.

56. B. Differential misclassification of the exposure in the setting of a case-control study is also called recall bias. In this example, case individuals who have pulmonary fibrosis are likely to know that they have a serious disease and may differentially recall their exposure to paint products based on this knowledge. Some non-differential misclassification of paint exposure may have also occurred in this study; however, this impact of this type of misclassification on the relative risk is predictable (bias toward the null). Misclassification of the study outcome is unlikely to have occurred due to the use of a highly accurate tool for determining the presence of pulmonary fibrosis. Undiagnosed pulmonary fibrosis is highly unlikely among the control group due to the absolute rarity of this disease and the lack of symptoms.
57. A. Recall bias is prevented in this example by ascertainment of paint exposure data collected prior to the occurrence of pulmonary fibrosis (outcome) and using an objective procedure. The company's record system may not perfectly classify the type, amount, and duration of previous paint exposure, potentially resulting in some non-differential misclassification of the exposure.
58. A. The study outcome, all-cause mortality, is likely to have been measured without any meaningful degree of error. However, the exposure, aspirin use, may have been misclassified, due to assessment at only a single point in time. If some aspirin users subsequently discontinued this treatment, or if some non-users initiated aspirin treatment, then these exposure groups would mix together, potentially diluting associations with the outcome. The relative risk observed in this study likely underestimated the association of aspirin use with mortality that would be obtained under gold standard measurements of aspirin use (e.g., annual blood tests for aspirin levels).
59. A. As suspected, some aspirin users discontinued this treatment during follow-up. Non-differential misclassification of the exposure remains the most likely pattern of error in this study because there is no compelling reason as to why errors in ascertaining aspirin use would differ systematically by subsequent mortality status.
60. C. In this example, apolipoprotein B (ApoB) levels measured by the automated laboratory platform are *consistently* 50% lower than those obtained by the gold standard reference procedure. Consequently, the same 20% of participants who were classified as having "elevated" ApoB levels by the automated platform would be classified as having elevated levels by the gold standard method.
61. D. A confounding characteristic is associated with the exposure, associated with the outcome, and does not reside on the causal pathway of association. Based on the data in Table 9.1, race, body mass index, intravenous drug use, and CD4 count are associated with the exposure, HIV disease. Of these characteristics, intravenous drug use and CD4 count are also associated with the outcome, CMV infection, based on the data in Table 9.2. However, CD4 count is likely to reside on the causal pathway between HIV disease and CMV infection: the plausible sequence of events is that HIV disease *leads* to a lower CD4 lymphocyte count, which then subsequently increases susceptibility of CMV infection.

62. C. There are a higher proportion of African Americans in the HIV-positive group compared with the HIV-negative group in this study, indicating an association of race with the exposure. On the other hand, there is no meaningful association between race and the outcome, CMV infection, indicated by a relative risk that is very close to 1.0 (Table 9.2). Race cannot plausibly reside on the causal pathway of association in this example.
63. A. These observational data demonstrate an association between the lowest category of treated systolic blood pressures and mortality. This association is subject to confounding by age, obesity, comorbidity, and any number of characteristics that might be linked with lower systolic blood pressures and higher rates of death. Given the (strong) likelihood of confounding in this example, it would be premature to conclude that overtreatment of hypertension was the cause of the observed higher mortality rate. Moreover, the study specifically evaluates the association of measured blood pressure with mortality but does not address the potential risks and benefits of specific blood pressure treatments. Many possible mechanisms may explain this observed association, including lower cerebral blood flow.
64. A. The causal diagram below organizes characteristics that are associated with the exposure, associated with the outcome, and suspected to reside on the causal pathway of association in this study.



Dietary fiber and regular exercise are associated with green tea consumption, but are not associated with pancreatic cancer in this study. Circulating concentrations of antioxidants and vasodilatory response are associated with green tea consumption and pancreatic cancer. However, given the background information motivating the study, antioxidants and vasodilatory response are presumed to result from green tea consumption; these characteristics are unlikely to influence a person's decision to drink green tea. Age is associated with green tea consumption and pancreatic cancer and age is not suspected to reside on the causal pathway of association.

65. E. Given the background information motivating this study, circulating concentrations of antioxidants are suspected to reside on the causal pathway between green tea consumption and pancreatic cancer. Therefore, this characteristic would *not* be considered to be a confounder in this study but instead a potential mechanism of association. Inference for a causal impact of green tea consumption on pancreatic cancer would *not* be enhanced by adjustment for circulating concentrations of antioxidants, whether by stratification, matching,

or restriction. Assuming that measured concentrations of antioxidants truly capture this characteristic, the relative risk adjusted for antioxidants can be interpreted as the association between green tea consumption and pancreatic cancer that is mediated through pathways other than anti-oxidants.

66. A. Stratification plus adjustment is a method to control for confounding. In this example, each stratum consists of only one sex thereby preventing confounding by sex within either stratum. The sex-adjusted relative risk can be calculated by weighing and combining the stratum-specific relative risks:

$$\text{Relative risk}_{\text{summary}} = (\text{relative risk}_{\text{women}} \times \text{weight}_{\text{women}}) + (\text{relative risk}_{\text{men}} \times \text{weight}_{\text{men}})$$

$$\text{Relative risk}_{\text{summary}} = (0.85 \times \text{weight}_{\text{women}}) + (0.85 \times \text{weight}_{\text{men}})$$

Although the number of people in each stratum is not provided, the weighted and combined relative risk must be 0.85 because the relative risk is 0.85 in each stratum. The sex-adjusted relative risk of 0.85 differs meaningfully from the unadjusted relative risk of 0.70, suggesting confounding by sex (a 10% or greater change in the size of the relative risk after adjustment can be used to suggest the presence of confounding).

67. D. Confounding-by-indication occurs when the indications for a medication or procedure are themselves confounding the association of interest. Randomly assigning large numbers of people to either a long-acting calcium channel blocker or a placebo will create similar distributions of characteristics across these groups and prevent confounding. Observed differences in stroke rates seen in randomized trials can be reasonably ascribed to the impact of calcium channel blocker treatment itself on this outcome. In contrast, differences in stroke outcomes seen in this observational study may be confounded by differences in the characteristics of users versus nonusers of long-acting calcium channel blockers.
68. B. Hypertension is the most likely confounding characteristic in this study because hypertension is the primary indication for prescribing a long-acting calcium channel blocker and because hypertension is a major risk factor for stroke. Restricting the entire study population (long-acting calcium channel blocker users and nonusers) to only people who have a diagnosis of hypertension would reduce the possibility of confounding by this characteristic, though residual confounding by subclinical hypertension would remain a possibility.
69. B. Matching, stratification plus adjustment, multivariate regression, and restriction are methods to control for confounding. P-values and 95% confidence intervals address the statistical significance of study findings, which is a different concept from confounding.
70. E. This case-control study matched case individuals, who had a diagnosis of atrial fibrillation, with control individuals who did not have an atrial fibrillation diagnosis by age (methods, page 79). Consequently, the cases and controls in this study had similar age distributions (Table 1 of the paper). Assuming that

age-matching was appropriately handled in the analysis, matching will control for confounding by age. A confounding characteristic requires linkage with both the exposure and the outcome. Breaking just one of these links is sufficient to control for the confounding influence of that characteristic. Age does *not* reside on the causal pathway between antihypertensive medication use and atrial fibrillation because it is impossible for these exposures to influence a person's age.

71. D. There is minimal change in the size of the association between total ACE inhibitor use and atrial fibrillation with adjustment (odds ratio changes from 0.81 to 0.79). Consulting the footnote of Table 3 in the paper, adjustment variables included smoking, body mass index, asthma, beta-lactam antibiotics, quinolones, macrolides, tetracyclines, corticosteroids, and beta-agonists. The lack of appreciable change in the odds ratio with adjustment suggests that these characteristics are *not* likely to be confounding the unadjusted association (though it is possible that some characteristics are confounding the observed association in opposing directions resulting in no net change). The adjusted association accounts for the characteristics included in the model, but not for the possibility of residual confounding.
72. B. In this large randomized trial, the proportion of anxiety disorder is likely to be balanced across the treatment groups at the start of the study. Confounding is unlikely in large randomized studies. The relative risk of persistent depressive symptoms associated with the new medication differs by the presence of anxiety disorder at baseline. The amount of difference in the relative risks is large and the confidence intervals for these relative risks do not overlap with one another. These findings suggest the presence of effect modification (on the relative risk scale).
73. C. This example demonstrates the application of stratification to address the concepts of confounding *and* effect modification. To evaluate the presence of confounding, stratum-specific associations are weighted and combined to yield an adjusted association. Without knowledge of the specific weights in this example, the sex-adjusted relative risk of cirrhosis, comparing hepatitis B-positive to hepatitis B-negative individuals, is somewhere between 2.3 and 2.9. Some sense of the weights can be gleaned from Table 9.1, which shows a roughly equal number of men and women in the study, suggesting that the sex-adjusted relative risk is probably close to 2.5. The lack of appreciable change in the size of the relative risk after adjustment for sex suggests that sex is *not* confounding the observed association between hepatitis B status and cirrhosis to any meaningful degree. To appraise the possibility of effect modification, the stratum-specific relative risks are compared *to each other*. In this example, the relative risks of 2.3 and 2.9 for women and men are roughly similar with widely overlapping confidence intervals, arguing against effect modification by sex. On the other hand, the relative risks are considerably different when stratified by the presence of heavy alcohol use, suggesting that this characteristic is modifying the association between hepatitis B status and cirrhosis.
74. D. The relative and attributable risks associated with vaccination are presented below.

	Incidence non-vaccinated	Incidence vaccinated	Relative risk	Attributable risk
Country 1	100	40	$100/40 = 0.40$	$100 - 40 = 60$ per 1000
Country 2	200	130	$200/130 = 0.65$	$200 - 130 = 70$ per 1000
Country 3	300	230	$230/300 = 0.77$	$300 - 230 = 70$ per 1000
Country 4	400	310	$310/400 = 0.78$	$400 - 310 = 90$ per 1000

Based on the difference in the size of the relative risks, the new vaccine appears to be most effective for preventing malaria in country 1. On the other hand, when attributable risk is used to quantify the size of these associations, the vaccine is associated with the greatest reduction in malaria incidence in country 4. Administration of the vaccine to country 4, where the background incidence of malaria is highest, would be expected to prevent the greatest number of malaria cases.

75. D. The study data demonstrate associations of NSAID use and ARB use with the incidence of AKI. However, the data as presented do not address whether there may be a synergistic impact of NSAID and ARB use on AKI incidence. The possibility of effect modification can be investigated by determining whether the association of NSAID use with AKI differs meaningfully according to the concomitant use versus nonuse of ARBs or conversely by determining whether the association of ARB use with AKI differs by the use versus nonuse of NSAIDs.
76. E. The cumulative incidence of ESRD is 30.8 per 10,000 among living donors compared with only 3.9 per 10,000 among a matched group of healthy non-donors. These differences represent a 7.9 times higher incidence of ESRD (although the absolute risk of developing ESRD is low in both groups). Could this association be confounded by age? Figure 2a of the paper demonstrates an association of older age with ESRD, the study outcome, among only the living donors group. However, healthy non-donors were matched to living donors by age, preventing linkage of age with the study exposure and thereby minimizing the possibility for confounding by age. To evaluate whether the size of the association between live kidney donation and ESRD differs by age, some measure of effect size, such as relative risk, must be provided across the age groups. The paper does not present such information (Fig. 2a presents data among only the living donors).
77. A. From the text of the results section (page 582), the incidences of ESRD among Black donors and Black non-donors are 74.7 and 23.9 per 10,000, respectively. Based on these incidences:

$$\text{Relative risk among Blacks} = (74.7) / (23.9) = 3.1$$

$$\text{Attributable risk among Blacks} = (74.7) - (23.9) = 50.8 \text{ per } 10,000$$

Among Hispanics, the incidences of ESRD are 32.6 and 6.7 per 10,000 for live donors and non-donors.

$$\text{Relative risk among Hispanics} = (32.6) / (6.7) = 4.9$$

$$\text{Attributable risk among Hispanics} = (32.6) - (6.7) = 25.9 \text{ per } 10,000$$

On the additive (attributable risk) scale, the association between live kidney donation and ESRD is higher comparing Black donors with Hispanic donors. The paper does *not* include data to address possible differences in the size of this association by age.

78. C. The new data present associations of live kidney donation with ESRD on the attributable risk scale. The size of the association is similar among the low versus high physical activity categories, suggesting no effect modification by this characteristic. The stratified data can also be used to evaluate confounding. Although the number of people in each physical activity group is not provided, the attributable risk of ESRD associated with live kidney donation must be somewhere between 12.3 and 13.8 per 10,000 after adjustment for physical activity status. This adjusted attributable risk differs considerably from the unadjusted attributable risk of 26.9 per 10,000, suggesting that physical activity *is* confounding the unadjusted association to a meaningful degree.
79. C. Screening tests are best applied to diseases that are serious, potentially treatable and include a preclinical phase that can be detected by a test. Melanoma is a serious, potentially treatable disease that is often preceded by a preclinical phase (mole, or dysplastic nevi). For influenza, the time between biological onset and clinical symptoms (preclinical period) is too short for a meaningful impact of screening. Mad cow disease is not treatable. Eczema and insomnia are not serious diseases and are unlikely to have preclinical phases that could be detected by a test before the onset of symptoms.
80. C. A useful method for solving screening and diagnosis problems is to create a 2×2 table, in which the columns represent gold standard evidence of the disease and the rows represent the test results. For ease of calculations, a total of 1000 people can be used as a round number for problem-solving.

	Disease	No disease	Total
Test positive	a	b	a + b
Test negative	c	d	c + d
Total	a + c	b + d	1000 total

The calculation of positive and negative predictive values always requires an estimate of the pretest probability of disease. In the context of screening, tested individuals are by definition free of signs or symptoms of the disease, providing no specific clues as to their pre-test disease likelihood. Consequently, for screening tests, the pretest probability of disease is estimated by the background prevalence of the disease among similar people in the population. In this example, the tested children have no symptoms of lead toxicity, so the 2% prevalence of this condition in their neighborhood serves as the pretest probability of disease.

	Lead toxicity	No lead toxicity	Total
Test positive	a	b	a + b
Test negative	c	d	c + d
Total	20 (2%)	980 (98%)	1000

Applying the 95% sensitivity and 98% specificity of the lead screening test to these data:

	Lead toxicity	No lead toxicity	Total
Test positive	$(20 \times 0.95) = 19$	196	215
Test negative	1	$(980 \times 0.80) = 784$	785
Total	20	980	1000

Among children who do *not* have actual lead toxicity, 196 will test positive (false-positive test).

81. E. Applying the identical sensitivity and specificity characteristics of the lead screening test to a second population in which the prevalence of disease is 50% yields the following data:

	Lead toxicity	No lead toxicity	Total
Test positive	$(500 \times 0.95) = 475$	100	575
Test negative	25	$(500 \times 0.80) = 400$	425
Total	500 (50%)	500 (50%)	1000

For a child who tests negative from this neighborhood, the probability of *not* having actual lead toxicity is $400/425 \times 100\% = 94\%$. This is the *negative predictive value* of the test

82. C. In this example, the prevalence of HIV disease is $(1 \text{ million} / 300 \text{ million} \times 100\%) = 0.33\%$, and the sensitivity and specificity of the test are 99.7% and 98.5%, respectively. If 300 million people are screened:

	HIV disease	No HIV disease	Total
Test positive	$(1,000,000 \times 0.997) = 997,000$	4,485,000	5,482,000
Test negative	3000	$(299,000,000 \times 0.985) = 294,515,000$	294,518,000
Total	1,000,000	299,000,000	300,000,000

Approximately 4.5 million people who do *not* have actual HIV disease will have a positive HIV antibody test. Given a very low prevalence of a disease, a specificity of 98.5% remains inadequate for suppressing false-positive tests.

83. B. From the above data, positive predictive value = $997,000 / 5,482,000 \times 100\% = 18.2\%$.

84. E. The same HIV antibody test is now used as a diagnostic test. In the setting of diagnosis, the pretest probability of disease is estimated from unique aspect of the clinical presentation. In this example, the pretest probability of disease is suspected to be 30%, based on the presenting symptoms and signs of this particular individual. A new 2×2 table is needed using a pre-test probability of 30% and the same sensitivity and specificity characteristics of the HIV antibody test. Given a total of 1000 hypothetical similar tested individuals:

	HIV disease	No HIV disease	Total
Test positive	$(300 \times 0.997) = 299$	10	309
Test negative	1	$(700 \times 0.985) = 690$	691
Total	300	700	1000

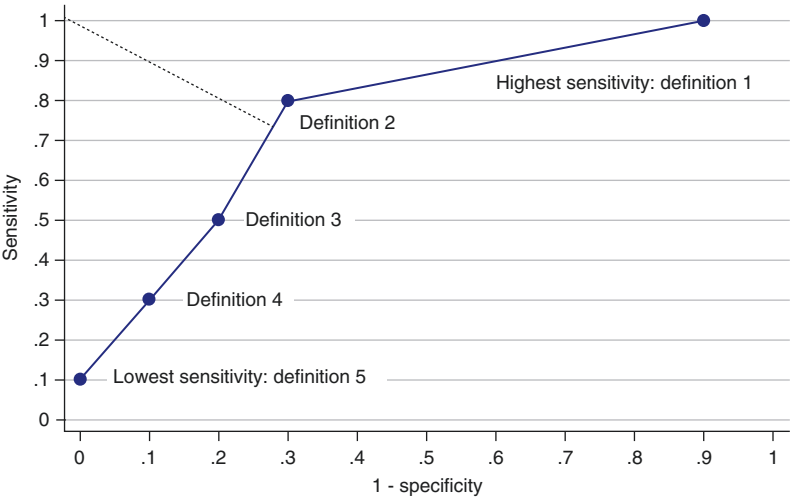
Positive predictive value = $299 / 309 \times 100\% = 96.8\%$

If this patient has a positive HIV antibody test, then the probability of having actual HIV disease is 96.8%.

85. B. For this 5-point diagnostic test, there are five mutually exclusive choices for defining a positive result. Each definition has a unique pair of sensitivity and specificity characteristics.

	Positive test	Negative test	Sensitivity	Specificity
Definition 1	Score ≥ 1	Score = 0	???	???
Definition 2	Score ≥ 2	Score < 2	???	???
Definition 3	Score ≥ 3	Score < 3	???	???
Definition 4	Score ≥ 4	Score < 4	???	???
Definition 5	Score = 5	Score < 5	???	???

The receiver operating characteristic (ROC) curve presents the sensitivity and specificity for *all* possible cut point choices. Unfortunately, ROC curves typically do not include labels to specify which specific points on the graph correspond to which specific cut points. However, it is possible to deduce this information from the graph. The lower left-hand corner of the ROC curve represents the cut point that yields the lowest sensitivity, which is defined as the proportion of people who test positive given the presence of disease. A cut point with low sensitivity implies that few people who have disease will test positive. For the new arthritis test, definition 5 has the lowest sensitivity because this stringent definition of a positive test will yield the fewest positive results. Moving along the ROC curve from lower left to upper right, the test sensitivity increases as specificity decreases. The upper right-hand corner of the ROC curve represents the cut point that yields the highest sensitivity and lowest specificity, as shown below.



Data from the ROC curve can then be used to complete the table of sensitivity and specificity characteristics.

	Positive test	Negative test	Sensitivity	Specificity
Definition 1	Score ≥1	Score = 0	0.9	0.1
Definition 2	Score ≥2	Score <2	0.8	0.7
Definition 3	Score ≥3	Score <3	0.5	0.8
Definition 4	Score ≥4	Score <4	0.3	0.9
Definition 5	Score = 5	Score <5	0.1	1.0

A diagonal line drawn from the upper left-hand corner to the ROC curve will identify the cut point with the mathematically optimal tradeoff between sensitivity and specificity. In this example, definition 2 lies closest to such a line. Note that the ideal cutoff value for a test depends not only on sensitivity and specificity but also on the prevalence of the disease in the tested population, the natural history of the disease if left untreated, the diagnostic procedures that follow a positive test, and the impact of treatment on people who test positive.

86. A. The negative predictive value of a test is defined as the probability of having no true disease among all people who test negative. Given a fixed prevalence of disease, the negative predictive value is directly related to the sensitivity of a test. Greater sensitivity will result in a greater negative predictive value due to a smaller number of false negatives. Definition 1 has the greatest sensitivity and highest negative predictive value.

	Disease	No disease	Total
Test positive	a	b	a + b
	True positive	False positive	
Test negative	c	d	c + d
	False negative	True negative	
	Test sensitivity	Test specificity	

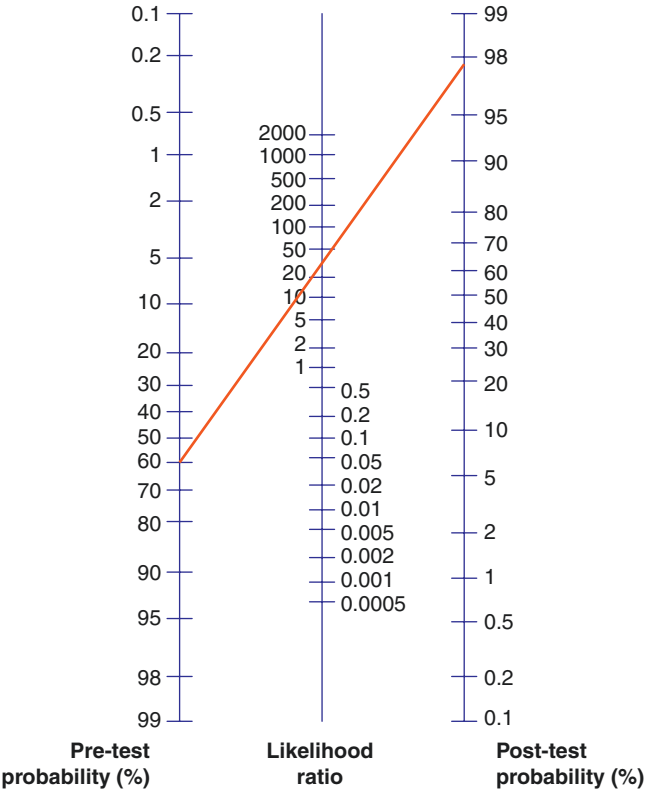
Negative predictive value = $d/(c + d)$

87. E. Data from the 100 test patients are used to determine the sensitivity and specificity of the new test. Among people with actual pancreatic cancer, 99% have a positive test; this is the sensitivity. Among people without actual pancreatic cancer, 98% have a negative test; this is the specificity. These values can then be applied to a population of 10,000 tested individuals among whom the prevalence of pancreatic cancer is 1%.

	Pancreatic cancer	No pancreatic cancer	Total
Test positive	$(100 \times 0.99) = 99$	198	297
Test negative	1	$(9900 \times 0.98) = 9702$	9703
Total	100	9900	10,000

A total of 297 patients will have a positive test (99 true positives and 198 false positives) and undergo subsequent computed tomography scanning.

88. A. Likelihood ratio nomograms provide an alternative method to 2×2 tables for calculating posttest probabilities from pre-test probabilities and the sensitivity and specificity of the test. Positive and negative likelihood ratios are derived from sensitivity and specificity. In this example, the pretest probability of disease is 60% and the positive likelihood ratio is 35. These values yield a posttest probability of approximately 98%, shown in the nomogram below (use a ruler or straight edge to estimate as closely as you can).



89. C. Validity (also called accuracy) describes the ability of a test to obtain the true value of the characteristic being measured. The validity of the new rhinovirus test refers to how closely the results of this test relate to actual rhinovirus infection, determined by a gold standard method.
90. A. Studies of screening are susceptible to biases, which are summarized in the table below.

Type of bias	Brief explanation
Confounding (referral bias)	People whose disease is detected by screening have different characteristics than those whose disease is detected at clinical presentation
Lead time bias	Extra follow-up time is preferentially added only to people whose disease is detected by screening
Length bias sampling	People whose disease is detected by screening have less aggressive forms of the disease
Overdiagnosis bias	Some people whose disease is detected by screening do not actually have the disease at all

For this question, survival time among people whose CKD is detected by screening begins when they have a positive test, which is relatively early in the course of the disease. In contrast, survival time among unscreened individuals begins much later, when CKD is detected by conventional clinical methods. This approach artificially adds survival time to the screened group, resulting in a lower mortality *rate* in this group and creating the false appearance of a survival benefit from screening.

91. B. The administration of a screening test at a single point in time will preferentially favor identification of the most prevalent variants of a particular disease. Prevalence is directly related to the duration of disease ($\text{Prevalence} = \text{Incidence} \times \text{Duration}$). Disease variants that progress slowly have a longer duration than variants that are rapidly progressive or quickly cured. In this example, screening is likely to preferentially identify the most slowly progressive types of kidney disease.
92. C. The point-of-care test in this example requires subjective interpretation. The researchers, who may be enthusiastic about the new test, may “overcall” some of the test strip results; for example, they may define a “dark orange” result as a positive test. Overdiagnosis will add healthy people, who do not have actual CKD, to the screen-positive group, resulting in better outcomes among this group.
93. C. The mean value of a distribution is sensitive to extreme values. The median and mode depend on the order of the observations in a distribution and are not impacted by extreme values. The geometric mean is less sensitive to extreme values than the standard (arithmetic) mean.
94. B. For normally distributed data, 67% of the observations will be located within one standard deviation of the mean value and 95% will be located within two standard deviations of the mean. In this example, approximately 95% of the study population will have a TSH level between $3.0 \pm (0.5 \times 2)$ or 3.0 ± 1 mIU/L. The remaining 5% will be evenly divided between TSH levels

of <2.0 mIU/L and TSH levels of >4.0 mIU/L if the data are normally distributed. Therefore, 2.5% of the study population, or 250 people, will have a TSH level >4.0 mIU/L.

95. B. Box plots present the median, interquartile range, and extreme observations for a given distribution they do *not* include the mean value. Distribution B has a slightly lower median value than distribution A but also has many extremely high values that are likely to have a large impact on the mean.
96. D. The figure demonstrates a reasonably normal distribution of serum sodium concentrations. Therefore, 95% of the data is expected to reside within two standard deviations of the mean value of 141 mEq/L. The lower 2.5th percentile is 138 mEq/L, which is 3.0 mEq/L from the mean value.

$$2 \times \text{standard deviation} = 3.0 \text{ mEq} / \text{L}$$

$$\text{Standard deviation} = 3.0 \text{ mEq} / \text{L} / 2 = 1.5 \text{ mEq} / \text{L}$$

97. D. The mode is the most frequently occurring value in a distribution. Based on the histogram, the most common coronary calcium score in the study is 0. The distribution of coronary calcium scores is *not* normal. The large rightward skew implies that the mean score will be much larger than the median score. The property that 95% of the data are located within two standard deviations of the mean applies to only normally (or roughly normal) distributed data and not to these coronary calcium scores.
98. E. Mortality can take on only two possible values: yes or no. The other characteristics are either continuous or categorical.
99. D. The correlation coefficient describes the tendency of larger values of one variable to match up with larger values of a second variable. A correlation coefficient of +1 indicates perfect positive agreement between two variables, a coefficient of -1 indicates perfect negative agreement, and a correlation coefficient of 0 indicates no agreement. The correlation coefficient of -0.15 for age and gait speed indicates that older age tends to track with lower gait speeds but that there is considerable variability in this association.
100. A. The white area on the quantile plot contains the middle three quintiles, or the 20th–80th percentile, of the distribution of serum phosphate levels for a given level of GFR. The edge between the white area and the upper gray-shaded area represents the 80th percentile of the distribution. This value is higher for a GFR value of 20 ml/min compared with a GFR value of 60 ml/min. The lower edge, which represents the 20th percentile of the distribution, is also higher for low GFRs. Among participants who have a GFR of 50 ml/min, the middle three quintiles, or 60% will have a serum phosphate level between 2.8 and 4.0 mg/dl. The mean phosphate level is similar for GFRs between 40 and 90 ml/min. The median value is not shown in the figure.
101. C. The null hypothesis *always* refers to the underlying population. The p-value is the probability of observing the study data, or more extreme study data, if the null hypothesis were true. If the two antibiotics actually had the same impact on end-

- of-study colony counts in the underlying population, then the chance of observing this 5.5 thousand colony count difference, or an even greater difference, is 4%.
102. D. The p-value is calculated from the sample size and the variance (or standard deviation) of the outcome. The table reports the sample size of 200 people in each group, but not the standard deviation.
103. A. This interpretation is correct. If this same study were repeated indefinitely, and a 95% confidence interval placed around the difference in mean end-of-study colony counts for each experiment, then 95% of these intervals will contain the actual difference in mean end-of-study colony counts in the underlying population (all women who are similar to those in the trial).
104. D. There is no way to determine whether this particular confidence interval of (0.2, 10.8) is one of the intervals that contains the actual difference in mean end-of-study colony counts in the underlying population. However, since 95% of all such intervals do contain the population value, we can be “95% confident” that the actual difference in mean end-of-study colony counts in the population is within this interval. Confidence intervals provide a range of estimated values in the underlying population.
105. B. Sample size and variability of the outcome influence the width of the 95% confidence interval. A larger sample size and smaller variation in the outcome will yield a narrower confidence interval, indicating greater certainty as to the true value of interest in the underlying population. It may be possible to reduce the variation in end-of-study colony counts by obtaining more reliable measurements of this characteristic (such as performing the colony counts in triplicate for each study participant).
106. C. Antibiotic One has a stronger impact on end-of-study colony counts in the randomized trial compared with the second “real-world” study. Chance should be considered as one possible explanation for this differential treatment effect; however, the 95% confidence intervals for end-of-study colony counts in the randomized trial and the clinic-based study do *not* overlap with one another, suggesting that chance is an unlikely explanation. The most likely reason for the differential treatment effects is the difference in the study populations and environments. The randomized trial was restricted to urinary tract infections that were caused by *E. coli* and included procedures to encourage compliance with antibiotic treatment.
107. A. The outcome of the study is a continuous variable (end-of-study colony count). The t-test is used to compare the mean value of a continuous variable between two different groups.
108. C. The outcome of this secondary analysis, recovery from infection, is a binary variable (yes versus no). Binary outcome data are often expressed as proportions: recovery occurred in $57/200 = 29\%$ of women assigned to antibiotic one and $70/200 = 35\%$ of women assigned to antibiotic two. The chi-square test is used to compare proportions between two different groups.
109. B. Statistically nonsignificant results may reflect the absence of a true association in the underlying population, errors in the design and conduct

of a study, or a type II error. Type II errors occur when a hypothesis test declares a study result to be statistically nonsignificant, yet a true difference exists in the population. Type II errors are often caused by inadequate study power. A study conducted in only 40 people is unlikely to detect a significant result for all but the most powerful treatments or risk factors.

110. D. Study power is influenced by sample size, the magnitude of the association in the underlying population, variability in the study outcome, and the threshold for declaring statistical significance. The best way to increase study power is to increase the sample size. Another approach is to optimize the precision of the study outcome variable, if possible. Studies of inconsistently measured outcomes will require more participants to demonstrate a significant association.
111. B. This is the third statistical test performed on the same data, raising the possibility that the p-value of 0.05 could represent a type I error (no true association between delirium and hospital stay in the underlying population yet the study obtains a statistically significant result). The replication study is inadequately powered to confirm or deny whether the results of the first study are likely to be true in the population. The first study of 40 people was barely able to declare statistical significance for a 2-day difference in hospital stay. Observing this same effect size in a study of only ten people will not reach statistical significance.
112. D. The linear regression equation for the unadjusted association between PM_{10} and FEV_1 is:

$$\text{Mean}(FEV_1) = \beta_0 + \beta_1 \times (PM_{10})$$

Based on the value reported in the table, the best fitting linear regression model for this association has a β_1 coefficient of -160 .

113. D. The linear regression equation for the unadjusted association between age and FEV_1 is:

$$\text{Mean}(FEV_1) = 4400 - 30 \times (\text{age})$$

A 40-year-old person in the study would have a predicted FEV_1 value of:

$$\text{Mean}(FEV_1) = 4400 - 30 \times (40) = 3200 \text{ ml}$$

The person of interest has an actual measured FEV_1 value of 3700 ml. The residual is the difference between the observed value and the value predicted by the regression model: $3700 - 3200 = 500$ ml.

114. C. Influential points are characterized by having an extreme X- and extreme Y-value. Observations with an extreme X-value, but an average Y-value (choice A), or observations with an average X-value, but an extreme Y-value

(choices B and D), typically do not exert a large influence on the slope of the regression line.

115. A. Fitted regression models are estimated from the observed study data. The unadjusted linear regression model relating age with FEV_1 was fit to a study population in which age ranged from 30 to 70 years; there are no observed data for people who are less than 30 years old. Consequently, there is no assurance that the results of the regression model will apply to people who are younger than 30 or older than 70.
116. A. Based on the adjusted association between PM_{10} exposure and FEV_1 , each 10 ug/m^3 greater PM_{10} value is associated with a 90-ml lower FEV_1 , on average, holding the other variables in the model constant. The negative sign preceding the FEV_1 coefficient in the table indicates an inverse association with PM_{10} . In linear regression, each unit difference in a predictor variable is related to an average difference in the outcome variable. Therefore, a 90 ml lower FEV_1 per 10 ug/m^3 greater PM_{10} is equivalent to a 9 ml lower FEV_1 per 1 ug/m^3 greater PM_{10} . Observational studies can adjust for suspected confounding characteristics, but residual confounding remains a concern, creating uncertainty as to the causal impact of the exposure of interest.
117. C. The null hypothesis is that there is no adjusted association between PM_{10} exposure and FEV_1 in the underlying population. Given this null hypothesis about the population, the chance of observing these experimental results or even more extreme results is 2%. This low p-value implies that there likely is an adjusted association between PM_{10} exposure and FEV_1 in the population of similar individuals.
118. C. A meaningful change in the coefficient of interest after adjustment implies that the adjustment variable(s) was confounding the observed association. In multiple regression, adjustment is performed by adding the potential confounding variables to the model. In this example, the coefficient for PM_{10} changes from -160 in the unadjusted model to -90 in the fully adjusted model. Such a meaningful change after adjustment for age, sex, current smoking, and height implies that one or a combination of these characteristics was likely to have been confounding the initial unadjusted association.
119. B. In log-link regression, the log of the mean value of a continuous outcome variable is modeled as a function of the predictor variables. In this example, the log (mean FEV_1) is on the left-hand side of the regression equation and the predictor variables are on the right-hand side.
120. B. Based on the adjusted model, each 10 ug/m^3 greater PM_{10} exposure is associated with a 2.4% lower FEV_1 value, holding the other variables in the model constant. Therefore, person two, whose PM_{10} exposure is 10 ug/m^3 higher than that of person one, would be predicted to have a 2.4% lower FEV_1 value. The 2.4% difference is expressed as the ratio of FEV_1 values comparing person one with person two.

$$FEV_1(\text{person one}) / FEV_1(\text{person two}) = 1.024$$

Given that person one has an FEV₁ value of 3900 ml:

$$3900 / \text{FEV}_1 (\text{person two}) = 1.024$$

Solving for the FEV₁ value of person two:

$$\text{FEV}_1 (\text{person two}) = 3900 \text{ ml} / 1.024 = 3809 \text{ ml}$$

121. D. A logistic regression model was used to obtain these results because they describe adjusted odds ratios:

$$\log(\text{odds asthma}) = \beta_0 + \beta_1 \times (\text{age}) + \beta_2 \times (\text{sex}) + \beta_3 \times (\text{smoking}) + \beta_4 \times (\text{PM}_{10})$$

In logistic regression, the antilog of each coefficient in the model represents the association between a one-unit difference in the predictor variable and the *odds ratio* of the outcome variable, holding the other variables in the model constant. In this example, each 10 ug/m³ higher PM₁₀ exposure is associated with a 40% greater odds of asthma, holding the other model variables constant.

122. C. The formal interpretation of the confidence interval of (1.1, 1.8) for this study is, “if this experiment were repeated an infinite number of times, and a 95% confidence interval constructed for the adjusted odds ratio of each experiment, then 95% of these intervals would contain the true adjusted odds ratio in the underlying population.” This confidence interval can also be interpreted (less formally) as, “there is ‘95% confidence’ that the true adjusted odds ratio of asthma associated with each 10 ug/m³ greater PM₁₀ exposure is between 1.1 and 1.8 in the underlying population.”
123. E. Kaplan-Meier plots describe the proportion of people who are alive and free of the study outcome over time. Differences in these Kaplan-Meier plots at 8 years can be determined by drawing a vertical line at the x-axis value of 8. The proportion of people who are alive and free of Parkinson disease after 8 years is about 55% in the positive family history group and 82% in the negative family history group, corresponding to a 27% difference. Kaplan-Meier plots present raw, unadjusted study data, without accounting for potential differences in participant characteristics, such as age. Differences in median survival can be calculated only if *all* of the comparison groups reach a survival probability of 0.5, which does not occur in the negative family history group. Censoring is defined as leaving a study for any reason *except* for incurring the specified study outcome. In this example, participants would be censored if they drop out, die, or the study ends, but not if they develop Parkinson disease, because this is the study outcome.
124. A. Hazard ratios very closely approximate relative risk for the time period of interest, which in this example is the entire duration of the study. Based on the data in the table, the unadjusted relative risk of Parkinson disease is two times higher comparing people with a family history of the disease to those without

a family history over the course of this study. Hazard ratios do not directly report on overall survival or death due to the disease. The procedure used to obtain the hazard ratios (the Cox proportional hazards model) accounts for potential differences in follow-up time among the exposure groups by performing repeated analyses each time the study outcome occurs.

125. C. In this somewhat difficult example, separate hazard ratios are presented for different intervals of follow-up time to demonstrate relative harm of the new synthetic graft in the first 30 postoperative days and then relative benefit thereafter. Summarizing the study data using only a single hazard ratio would have concealed this important distinction. A hazard ratio of 3.0 during the first 30 postoperative days implies that there are three times as many deaths in the new synthetic graft group than the standard graft group during this period. Given a total of 160 deaths during this period, we can infer that 120 deaths occurred in the new synthetic graft group and 40 deaths occurred in the standard graft group. Given 1000 participants in each group at the start of the trial and minimal dropout:

Postoperative day	Survival in new synthetic graft group	Survival in standard graft group
0	1.0	1.0
30	120 deaths	40 deaths
	$880/1000 = 0.88$	$960/1000 = 0.96$

During the next period (days 31–60), there were a total of 240 deaths with a corresponding hazard ratio of 0.5. This hazard ratio implies that there were 80 deaths in the new synthetic graft group and 160 deaths in the standard graft group during this period (assuming no dropouts).

Postoperative day	Survival in new synthetic graft group	Survival in standard graft group
0	1.0	1.0
30	$880/1000 = 0.88$	$960/1000 = 0.96$
60	$800/1000 = 0.80$	$800/1000 = 0.80$

Overall survival is equal between the two groups after 60 postoperative days.