Veterinary Clinical Epidemiology

From Patient to Population

Fourth Edition

Ronald D Smith

CRC Press
Taylor & Francis Group

Veterinary Clinical Epidemiology

From Patient to Population

Fourth Edition

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Veterinary Clinical Epidemiology

From Patient to Population

Fourth Edition

Ronald D. Smith

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper

International Standard Book Number-13: 978-1-138-39242-7 (Paperback)
978-1-138-39298-4 (Hardback)

Contents

List of Examples

1. INTRODUCTION

2. DEFINING THE LIMITS OF NORMALITY

3. EVALUATION OF DIAGNOSTIC TESTS

4. USE OF DIAGNOSTIC TESTS

5. MEASURING THE COMMONNESS OF DISEASE

6. RISK ASSESSMENT AND PREVENTION

7. MEASURING AND COMMUNICATING PROGNOSES

8. DESIGN AND EVALUATION OF CLINICAL TRIALS

9. STATISTICAL SIGNIFICANCE

11. MEASURING AND EXPRESSING OCCURRENCE

12. ESTABLISHING CAUSE

13. SOURCE AND TRANSMISSION OF DISEASE AGENTS

14. THE COST OF DISEASE

Preface to the Fourth Edition

Veterinary education and practice have evolved considerably since publication of the third edition of this book in 2005. The veterinary knowledge base has expanded; specialized care and referrals are more commonplace; new diagnostic techniques, medications, treatments, immunization, and pain management protocols have emerged; and patient demographics have changed. In response, veterinary education is becoming more specialized, with many schools offering a tracking curriculum in a practice theme or discipline area, with a focus on producing graduates adept at problem solving, critical thinking, communicating, and self-directed learning. The discipline of clinical epidemiology is especially suited for this environment.

Clinical epidemiology may be defined as the research discipline concerned with the application of epidemiologic methods to questions directly relevant to the practice of medicine at the individual or population level. Clinical epidemiology focuses on the sorts of questions asked in the practice environment. It provides the tools to help practitioners apply their own experiences, the experiences of others, and the medical literature to medical decision-making. It is a basic science for the clinician.

This book is intended to introduce veterinary students, recent graduates, residents, and practitioners to epidemiological concepts and methods in a clinical context, and improve the reader's ability to critically evaluate medical claims and find evidence-based solutions to clinical questions. Emphasis is placed on proficiencies needed by graduates upon entry into clinical practice. The chapter sequence retains its problem-oriented approach to veterinary practice. Content has been updated to reflect new methods and concepts, with expanded coverage of risk, statistical and economic analyses, and disease surveillance. More than 60 examples of clinical research drawn from the international veterinary practice literature are presented as structured abstracts, a format that facilitates the communication of epidemiological methods and findings. Follow-up questions invite the reader to participate in the analysis of results. Full-text versions of more than half of these abstracts and more than 40% of the book's 174 literature citations are freely available online providing an opportunity for more in-depth exploration of these reports by the reader. If you have the electronic version of this book, you can click through the links to these. For those who have the print version of the book, "clickable" versions of the links are available via the "Downloads" tab at https://www.crcpress.com/9781138392427.

The ready availability of online veterinary medical information has redefined the veterinarian-client-patient relationship. Veterinarians are increasingly playing the dual roles of users of this information and interpreters for their clients. This book is intended to support practitioners in this role.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Acknowledgments

I want to thank Ms. Alice Oven, Senior Editor, Veterinary Medicine, CRC Press/Taylor & Francis Group, for her continual guidance and support during the course of preparing this revision. This revision was greatly facilitated by the ready access to searchable online resources such as the Veterinary Information Network (VIN) and full-text versions of veterinary literature provided through the Library of the University of Illinois at Urbana-Champaign. This edition of the book relies on content from a broad collection of international journals, and ready access to full-text versions of these journals was invaluable. I must also recognize the contributions of the many fine veterinary practitioners and researchers whose works are cited profusely throughout the text.

The task of preparing the fourth edition of this book was made easier by the continued understanding and support of my wife, Lupe. Our daughter, Veronica, was instrumental in the design of the book's cover, and our son, Ronald, provided valuable insight into the world of online databases and cloud computing.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

About the Cover

Richard Hess, "Peaceable Kingdom" from Robert Funk Fine Art

The cover image for this 4th Edition is a reproduction of an acrylic-on-canvas work by Richard Hess (1934–1991) titled *Peaceable Kingdom, Zebra, Buffalo, Lion, Giraffe, Elephant, Monkey, Tiger, Gorilla, Kangaroo etc*, ca. 1980. The image depicts a selection of the multitude of species, including humans, about whose health veterinarians must be concerned. It also conveys a sense that "We're all in this together," vulnerable to the same kinds of risks, diseases, diagnostic challenges, and outcomes, some of which (the zoonoses) are shared among us. Indeed, more than 800 pathogens are shared by humans and other animals, many of which are described as emerging diseases. These concepts are integral to the current "One Health" movement that has become increasingly important in medical education. Finally, the variety of animals staring out at us from the cover convey the responsibility of us all to protect species diversity on our planet. They seem to be asking, "What's going to become of us?" Climate change, habitat loss, and epidemics threaten the survivability of most nondomestic species on the planet. Several of the examples in this book illustrate the contribution that veterinarians have made toward reducing the impact of disease on wildlife populations.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Author

Ronald D. Smith, **DVM, PhD**, is a Professor Emeritus of Epidemiology and Preventive Medicine in the College of Veterinary Medicine at the University of Illinois. He received his DVM from Michigan State University in 1967 and his MS and PhD degrees in veterinary medical science from the University of Illinois.

From 1967 to 1970, Dr. Smith served as a U.S. Peace Corps veterinarian in Ecuador, working primarily on preventive disease programs for cattle, swine, and horses. He joined the faculty of the University of Illinois College of Veterinary Medicine in 1974. Dr. Smith's research interests have focused on the diagnosis, epidemiology, and control of vector-borne blood diseases of animals and veterinary medical informatics.

Dr. Smith has undertaken numerous consultancies throughout Central and South America on behalf of IICA, FAO, and IAEA in support of tickborne disease control programs. He has taught professional and graduate courses on veterinary epidemiology, food safety and public health, and medical informatics. He has presented numerous invited papers at international conferences and is principal author or coauthor of more than 90 scientific publications and scholarly works.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1 Introduction

1.1 DEFINITIONS

Over the years, there have been many definitions of epidemiology. Some examples follow:

A. "…the study of the health status of populations…" (Schwabe et al., 1977)
B. "…Epidemiology is nothing more than ecology with a medical and mathematical flavor." (Norman D. Levine, 1990, personal communication)
C. "The branch of medicine that deals with the study of the causes, distribution, and control of disease in populations." (*American Heritage Medical Dictionary,* 2007)
D. "The study of the determinants of disease events in populations." (*Mosby's Medical Dictionary,* 2009)
E. "Epidemiology is the study of the distribution and determinants of health-related states or events (including disease), and the application of this study to the control of diseases and other health problems." (World Health Organization, 2019, https://www.who.int/topics/epidemiology/en/)

Common threads in the above definitions are revealed if we consider their origin (Wikipedia contributors, 2019, https://en.wikipedia.org/wiki/Epidemiology). The term *epidemiology* derives from three Greek words: *epi* ("about" or "upon"), *demos* ("populace" or "people of districts"), *logos* ("word," thus science or theory). The term *epizootiology* is sometimes used in reference to comparable studies in animal populations. The distinction is useful when one wishes to describe the state of disease in human or animal populations specifically, particularly when discussing zoonotic disease. For most purposes, however, epidemiology is understood to refer to all animal populations, human and otherwise. Likewise, to avoid confusion, it is preferable to use the term *epidemic* in lieu of *epizootic*, and *endemic* in lieu of *enzootic* wherever possible. Thus, **a simple definition of epidemiology** that captures the spirit of earlier definitions and reflects the emphasis of this book is "*…the research discipline concerned with the distribution and determinants of disease in populations*" (Fletcher et al., 1982).

This definition alone does not appear to provide sufficient grounds for creating a separate discipline. After all, laboratory researchers study disease in populations of animals, populations that may comprise hundreds or thousands of individuals. Furthermore, laboratory researchers address the same sorts of questions as do epidemiologists—questions such as the cause, clinical signs, diagnosis, treatment, outcome, and prevention of disease. An important distinction, however, is that **epidemiologists study disease in its natural habitat, away from the controlled environment of the laboratory**. Epidemiology deals with naturally or spontaneously occurring, rather than experimentally induced, conditions.

The foregoing definitions imply that epidemiology is concerned with the population rather than the individual. To a certain extent this is true. However, an understanding of health and disease in populations is fundamental to medical decision-making in the individual.

The discipline of epidemiology is a critical component of the **One Health** initiative. One Health focuses on delivering collaborative, multidisciplinary solutions to complex problems at the animal, human, and environmental interface. This approach brings together the strengths of multiple health science professionals including veterinarians, physicians, public health professionals,

epidemiologists, ecologists, economists, social scientists, toxicologists, and others—working locally, nationally, and globally—to attain optimal health for people, domestic animals, wildlife, plants, and our environment. Examples of problems that are the focus of the One Health approach are zoonotic and emerging diseases, food safety, antimicrobial resistance, disaster preparedness, and disease surveillance and control in domestic animals and wildlife. One Health embraces the idea that complex problems at the human-animal-environmental interface can best be solved through multidisciplinary communication, cooperation, and collaboration across disciplines and sectors.

1.2 EPIDEMIOLOGIC APPROACHES

Epidemiology has its roots in disease surveillance and outbreak investigation. Many consider that epidemiology was "born" during the cholera investigations conducted by John Snow in London in the mid 1800s. However, examples of outbreak investigation can be documented as far back as the Greek and Roman eras (Morens, 2003). Over the years, a number of epidemiologic disciplines and associated methodologies have emerged. These categories are somewhat arbitrary but illustrate some of the ways in which epidemiology contributes to veterinary and human medicine.

1.2.1 Quantitative Epidemiology

Quantitative epidemiology strives to quantify the distribution of diseases and associated factors in terms of individuals, place, and time and explore potentially causal associations. Quantitative epidemiology is practiced at two levels: **descriptive** and **analytic**. Descriptive statistics may be expressed as numerator data (number of individuals), proportions, rates, or in terms of central tendency and dispersion. Data-gathering methods include sampling and diagnostic techniques for detecting the presence of disease, surveillance techniques for monitoring disease activity, and record-keeping systems. The submission of patient encounter data from U.S. veterinary medical teaching hospitals to the Veterinary Medical Database (VMDB) is an example of a descriptive, data-gathering technique. VMDB is a repository of data on patients from more than 7 million hospital records submitted by 26 universities since 1964. The adoption of electronic veterinary medical records (EVMRs) in independent companion animal practices is another example of the passive capture of patient-encounter data that can potentially be used to improve patient care and detect emerging or exotic animal diseases. A 2010 survey of independent small animal practices in Massachusetts revealed that EVMRs were used alone or together with paper records by 66 of 82 (80.5%) responders (Krone et al., 2014). Veterinarians in paper record—only practices indicated that reluctance to change, anticipated technological problems, time constraints, and cost were barriers to EVMR use. Additionally, determining the accuracy of the EVMR is vital to the progress of practice-based research (Robinson et al., 2015).

Other examples of the passive capture of animal disease data are the monitoring and surveillance activities of the USDA's National Animal Health Monitoring System (NAHMS) and the World Organization for Animal Health's (OIE) World Animal Health Information System (WAHIS) for monitoring priority diseases of terrestrial and aquatic animals worldwide. Results are expressed as descriptive statistics. Historical surveillance data provide an especially useful point of reference for documenting changes in disease frequency from such diverse causes as new and emerging diseases or adverse reactions to new pharmaceuticals or vaccines.

Analytic epidemiology goes beyond the purely descriptive process to draw statistical inferences about disease occurrence and possible causal associations. Techniques employed include univariable and multivariable regression, clustered and spatial data analysis, survival analysis, decision analysis, risk analysis, mathematical modeling, and a variety of statistical tests of significance. These techniques may be used to help distinguish true causal relationships from those simply due to bias, confounding, or chance, a problem inherent to epidemiologic research.

1.2.2 Ecological Epidemiology (Medical Ecology)

Ecological epidemiology focuses on understanding factors that affect transmission and maintenance of disease agents in the environment. These factors are sometimes referred to as the ***agent-host-environment triad***. Ecological epidemiology provides the scientific foundation for past and present disease eradication programs. The successful eradication programs for Texas cattle fever (bovine babesiosis) and screwworm (*Cochliomyia hominivorax*) were conceived based on knowledge of the natural history of the respective diseases. Traditionally, ecological epidemiology has focused on the life cycle, or natural history, of disease. The integration of molecular biology into traditional epidemiologic research, e.g., "**molecular epidemiology**," has provided new tools for studying disease occurrence at the molecular level.

1.2.3 Etiologic Epidemiology

Etiologic epidemiology is primarily concerned with exploring causal relationships for diseases of undetermined origin. Other terms that have been used to describe this activity are "medical detection," "shoe leather," and "field" epidemiology. One of the principal activities in this category is outbreak investigation. Investigation into the cause(s) of food-borne disease outbreaks is a classic example of etiologic epidemiology. A variety of sophisticated analytic techniques have been developed to help assess the relative importance of multiple causes of disease.

1.2.4 Herd Health/Preventive Medicine

Herd health/preventive medicine uses information from any or all of the sources mentioned previously to design optimal management, control, or preventive strategies. Sometimes this requires a formal risk analysis to determine the true impact of presumed risk factors. Economic considerations are often the basis for determining which strategy is most effective. The most effective strategy may not be the one that results in the lowest incidence of disease, but rather the one that results in the greatest profit. Veterinary practitioners must learn to think in these terms if they are to interact effectively with producers.

1.2.5 Clinical Epidemiology

Clinical epidemiology may be defined as ***the research discipline concerned with applying epidemiologic methods to questions directly relevant to the practice of medicine at the individual or herd/flock level***. The sorts of questions asked in the practice of medicine are listed in Table 1.1. The answers to these questions are of immediate relevance to disease diagnosis, risk appraisal, prognosis, and treatment. Study designs may be observational or experimental. Observational studies represent a formal approach to the inductive process by which practitioners turn their practical observations into experience. Experimental studies (clinical trials) evaluate the relative merits of various interventions such as therapeutic, surgical, or preventive approaches to a particular disease syndrome. Clinical epidemiology provides the tools to help practitioners apply their own experiences, the experiences of others, and the medical literature to medical decision-making.

> ***Epidemiologists study disease in its natural habitat, away from the controlled environment of the laboratory. Clinical epidemiology focuses on the sorts of questions asked in the practice of medicine.***

TABLE 1.1
Clinical Issues and Questions in Veterinary Practice

Issue	Question
Normality/Abnormality (Ch. 2)	What are the limits of normality?
	What abnormalities are associated with having a disease?
Diagnosis (Ch. 3, 4)	How accurate are the diagnostic tests or strategies used to find a disease?
Frequency/Occurrence (Ch. 5, 11)	What is the case definition for a disease; how common are each of the findings?
	What are the host, spatial, and temporal distribution of the disease?
Risk/Prevention (Ch. 6, 14)	What factors are associated with the likelihood of contracting disease?
Prognosis (Ch. 7)	What are the consequences of having a disease?
	What factors are associated with an increased or decreased likelihood of recovering from disease?
Treatment/Control (Ch. 8, 14)	How effective is a therapeutic strategy and how does it change the future course of a disease?
	How can the risk and rate of spread of the disease be reduced? How useful are the available tools for diagnosis, treatment, control, and prevention?
Chance (Ch. 9)	How confident can we be in clinical research findings?
Cause (Ch. 10, 12)	What is the etiologic agent? What is its life cycle? What characteristics contribute to its pathogenicity and virulence?
	What factors determine the susceptibility or resistance of individuals to the disease?
	What conditions predispose populations to outbreaks?
Source/Transmission (Ch. 13)	What is the source and reservoir mechanism of the causative agent? What are the periods of communicability?
	How is the agent spread from infected to susceptible individuals? What is the route of infection?
Cost (Ch. 14)	What is the impact of a disease in personal and economic terms?

Source: Adapted from Table 1.1 in Fletcher RH et al. *Clinical Epidemiology—The Essentials*, 1st ed. Baltimore: Williams and Wilkins; 1982. With permission.

1.3 APPLICATIONS OF EPIDEMIOLOGY IN VETERINARY PRACTICE

Epidemiology has been described as a basic science for clinical medicine (Sackett et al., 1991). Epidemiologic studies are often the only way of exploring clinical issues such as the accuracy of diagnostic tests, risk factor identification, cause of diseases of multiple or uncertain etiology, and disease prognosis with and without treatment. They also provide a means for studying rare conditions or complications of disease that would be difficult to induce experimentally. The practitioner's own patients represent an important source of epidemiologic data. The cumulative clinical experience captured in a patient database can be used to evaluate and improve patient care. Epidemiology also provides the tools for critical evaluation of medical claims. Bias, methodological errors, invalid assumptions, and chance can lead to erroneous conclusions from clinical studies. As one author put it: "…s*cience is the currency of medicine and the standard by which therapeutic claims are judged*" (Ramey, 2003a).

The relationship between epidemiology and clinical medicine has been formalized in the practice of ***evidence-based medicine*** (EBM), *the judicious integration of best research evidence with the patient's values to make decisions about medical care* (NLM, 2018). (McKenzie, 2012).

EBM consists of the following five steps (Sackett et al., 1997):

1. At each stage of the case workup, identify one or more clinically important information needs and convert them into answerable questions.
2. Track down, with maximum efficiency, the best evidence with which to answer the above questions.
3. Summarize and critically appraise the evidence found for scientific validity and applicability.
4. Apply the results of this appraisal to patient care.
5. Evaluate your performance at answering the question(s).

Although it may not be necessary for a practitioner to follow these steps for every case, most would probably agree that medical claims should be supported by evidence derived from patient experience.

EXAMPLE 1.1: ARE THERE TOO MANY ANESTHETIC DEATHS IN VETERINARY PRACTICE? WHAT IS THE ANESTHETIC DEATH RATE IN SMALL ANIMAL PRACTICE? ARE THERE ANY IDENTIFIABLE RISK FACTORS?

Background: The risk of anesthetic death has been studied in dogs and cats by teams in several countries for several decades and has been reported to be approximately 10 times that for human anesthesia patients.

Objectives: Bille et al. (2012, 2014) used an **evidence-based medicine** approach to develop a logistic regression model for the identification of risk factors for anesthetic death among dogs and cats, defined as death occurring during or before full recovery from anesthesia, and apply their findings to cohorts of patients seen in their private practice.

Study Design: Cohort study.

Methods: All dogs and cats that underwent general anesthesia at the Centre Hospitalier Vétérinaire des Cordeliers, Meaux, France were included in the study. Animals that were sedated or anesthetized in order to be euthanized were not included. During **study period 1**, a total of 3546 animals undergoing general anesthesia were used to calculate death rates and develop a logistic regression model to identify potential risk factors. During **study period 2**, three recommendations, relating to improving physical status and anesthetic/analgesic regimen, were implemented in 2685 patients and death rates compared with those of period 1.

Results: The overall death rate during period 1 was 1.35% and during period 2 was 0.8%. For sick animals (ASA status 3 and over), the overall death rate was 4.8% during period 1 and 2.2% during period 2. This represented a significant decrease in death rate in period 2 ($p = 0.002$). In period 2, the main factors associated with an increased likelihood of anesthetic death were poor health status (ASA physical status classification) and old age (Table 1.2). Species, gender, anesthetic regimen, and the nature and urgency of the procedure were not associated with risk.

Conclusions and Significance: The authors conclude that by following evidence-based recommendations they were able to significantly reduce anesthetic mortality. They caution that other factors not monitored in this study may also have influenced outcomes.

FOLLOW-UP QUESTION 1.1

How many of the comparisons in Table 1.2 show a statistically significant increase in the risk of anesthetic death ($p < 0.05$)? (*Hint:* See Chapter 6, "Risk Assessment and Prevention for a clue.") See Answer 1.1 at the end of this chapter.

TABLE 1.2
Logistic Regression Models of the Odds of Anesthetic Death in Animals Undergoing Anesthesia during Period 2

	OR	95% CI	*p*-Value
ASA category[a]			
3	–	–	–
4	6	1.8–19.6	0.00
5	81.6	20.8–321	0.00
Species			
Dog	–	–	–
Cat	2.2	0.7–6.4	0.16
Type of procedure			
Examination	–	–	–
Surgery	2.4	0.8–7.2	0.11
Geriatric[b]			
No	–	–	–
Yes	3.6	1.2–10.9	0.03

Source: Bille C et al. *Vet Anaesth Analg* 2014;41:249–58. With permission.
Abbreviations: Premed, Premedication; ASA, Association of American Anesthesiologists.
[a] The OR represents a change from ASA 3 to ASA 4 or ASA 3 to ASA 5.
[b] The OR represents a change from non-geriatric to geriatric.

The appropriate use of **complementary and alternative veterinary medicine** (CAVM) provides an opportunity to appreciate the implications of evidence-based medicine. Although CAVM options have been promoted for preventing or treating a broad range of animal ailments, there is a paucity of clinical studies (evidence sources) upon which to evaluate their efficacy and effectiveness (Ramey, 2003b, McKenzie, 2012). It is therefore difficult for CAVM-based medical claims to meet the criteria defined in steps 2 and 3 above. This does not mean that CAVM-based approaches don't work. It simply means that the choice of any therapeutic modality should be based on a critical evaluation of its scientific basis and evidence of a favorable outcome. If a client insists on adopting an alternative modality for which little or no clinical evidence exists, the practitioner should offer to assist in monitoring and evaluating the response in a critical but sympathetic way (Rollin, 2002).

Meta-analysis is another tool for converting clinical experience into practice guidelines. Meta-analysis consists of a quantitative systematic review of pooled results from several studies that investigate the same clinical issue. The end result is a summary of the evidence for and against current recommendations. It provides a bird's-eye view of research findings by analyzing numerous clinical research reports focusing on the same clinical question.

EXAMPLE 1.2: HOW EFFECTIVE ARE TREATMENTS FOR FELINE URINE SPRAYING?

Background: Feline urine spraying inside the home is a common behavioral problem prompting owners to seek veterinary advice. Individual trials relating to a variety of interventions have produced variable results, and to date no consensus on the value of different treatments has emerged.

Objectives: Mills et al. (2011) sought to synthesize the current data from published clinical trials that evaluate treatments for feline urine spraying to discern the influence that non-behavioral intervention methods have on the incidence of either the cessation of urine spraying or its reduction.

Study Design: Meta-analysis of published, peer-reviewed clinical trials.

Methods: Inclusion and exclusion criteria for study selection were predefined and methodological quality was assessed by two independent reviewers. Ten studies in nine publications that either evaluated pharmacotherapy or pheromonatherapy (the use of a synthetic analogue of the F3 facial fraction in the cat) were suitable for analysis. All cats were studied in their home environment.

Results: There was a significant ($p < 0.0001$) association between the use of any intervention and the number of cats that ceased or reduced urine spraying by at least 90%. Analysis by intervention type indicated that fluoxetine, clomipramine, and pheromonatherapy may each assist in managing urine spraying beyond a placebo-based intervention. The sustained use of fluoxetine had the largest reported effect according to two relatively small studies. Further evaluation of this treatment is required to establish whether the results can be replicated with larger sample sizes.

Conclusions and Significance: There is good evidence that both pharmacological and pheromonal interventions provide added value for the reduction of urine spraying in the cat. It is worth noting that the most comprehensive treatment program described, i.e., the one involving a triple-line intervention consisting of psychopharmacology (fluoxetine), environmental modification, and a cleaning regime for the longest period of time, appears to be the most effective treatment documented to date. The authors suggest that future research into treatment efficacy for this problem uses the benchmark standard of randomized, controlled trials lasting for at least 8 weeks, with the outcome criteria of cessation of feline urine spraying or reduction by at least 90%.

FOLLOW-UP QUESTION 1.2

The authors suggest that randomized controlled trials be conducted to confirm the efficacy of the above treatments for feline urinary spraying. What sources of study bias might be addressed by this approach? (*Hint:* See Chapter 8, "Design and Evaluation of Clinical Trials for a clue.") See Answer 1.2 at the end of this chapter.

1.4 OBJECTIVES

This text is intended to give you a working knowledge of veterinary epidemiology. Specifically, it (1) shows you how epidemiologic data are used in medical decision-making, (2) familiarizes you with epidemiologic study designs that allow valid conclusions to be drawn while controlling for sampling bias and chance, and (3) helps you learn to critically review and extract useful information from the medical literature. This is not intended to be a methods book. Readers can consult the cited articles from which examples were taken to learn more about particular methods.

1.4.1 Development of Medical Decision-Making Skills

Medical curricula, both human and veterinary, tend to focus on the mechanisms of disease in the individual through the study of anatomy, physiology, microbiology, immunology, and other basic sciences. This fosters the belief that the correct diagnosis and treatment of disease depends entirely on learning the detailed processes of disease in the individual. In medical practice

we deal with uncertainties, expressed as probabilities or risk. Each member of a population affected by the same disease agent may display a unique combination of signs. The frequency distribution of signs exhibited by the affected population will influence the accuracy of your diagnoses, prognoses, and treatments. An understanding of this variability can help you choose and interpret diagnostic tests and make clinical decisions. A practical problem resulting from disease variability is that of "case definition," the starting point for determining the effectiveness of new therapeutic regimens.

EXAMPLE 1.3: WHAT ARE THE CLINICOPATHOLOGICAL FINDINGS IN DOGS NATURALLY INFECTED WITH *LEISHMANIA INFANTUM*, AND HOW CAN THEY BE USED DIAGNOSTICALLY?

Two properties of diagnostic tests that affect their performance are sensitivity and specificity. Sensitivity data may not be recognized as such when used to describe clinical findings in patients. Canine leishmaniosis (CanL) due to *Leishmania infantum* is a zoonotic disease transmitted by a phlebotomine sand fly vector. Clinical manifestations in dogs are highly variable, ranging from subclinical forms to overt severe fatal disease. Table 1.3 summarizes clinical findings among 51 dogs in which a diagnosis of *L. infantum* infection was made based on serology and/or observation of the organism in tissue specimens (Meléndez-Lazo et al., 2018).

FOLLOW-UP QUESTION 1.3

Which of the findings ranked in Table 1.3 provides the best criteria for ruling out a diagnosis of leishmaniasis? (*Hint:* See Chapter 4, "Use of Diagnostic Tests for a clue.") See Answer 1.3 at the end of this chapter.

TABLE 1.3
Frequency of Clinical Signs in Dogs Infected with *Leishmania infantum*

Clinical Signs	Number of Dogs (%)
Cutaneous lesions	40 (78.4)
Lymphadenomegaly	33 (64.7)
Local	24 (47.1)
Generalized	9 (17.6)
Weight loss	24 (47.1)
Ocular lesions	13 (25.5)
Fever	11 (21.6)
Vomiting, diarrhea	10 (19.6)
Lameness	6 (11.8)
Joint pain	3 (5.9)
Other causes	3 (5.9)
Polyuria and polydypsia	5 (9.8)
Epistaxis	4 (7.8)
Muscle atrophy	4 (7.8)
Splenomegaly	3 (5.9)

Source: Meléndez-Lazo L et al. *Res Vet Sci* 2018;117:18–27. With permission.

1.4.2 Learn Epidemiologic Methodology and How to Analyze and Present Data

The science of epidemiology evolved from the need to study naturally occurring health and disease in populations. The study of health and disease away from the controlled environment of the laboratory increases the likelihood that bias, confounding, and chance will influence our findings. The tools of epidemiology include a variety of techniques for collecting, analyzing, and interpreting data. They enable one to draw accurate conclusions about populations by controlling for bias, confounding variables, and random error. Summary presentation of data as tables or graphs can help clarify relationships and trends. Computer simulation models of disease dynamics is an increasingly important method for incorporating field data into predictive models.

A familiarity with descriptive and inferential statistics should be a prerequisite for veterinarians, who are continually faced with the risk of misdiagnosing a case. The design of animal disease surveillance programs is influenced by sampling and detection statistics. Private practitioners may be asked to participate in state and federal regulatory efforts and must understand their scientific basis. Veterinarians that have been accredited through the USDA-APHS National Veterinary Accreditation Program (https://www.aphis.usda.gov/aphis/ourfocus/animalhealth/nvap) play a critical role in protecting animal and human health through animal testing and signing health certificates for interstate movement. Current U.S. control or eradication efforts focus on brucellosis, Johne's disease, pseudorabies, tuberculosis, transmissible spongiform encephalopathies, scrapie, bovine spongiform encephalopathy, and chronic wasting disease.

EXAMPLE 1.4: WHAT IS THE BEST STRATEGY FOR CONTROLLING CHRONIC WASTING DISEASE (CWD) IN FREE-RANGING DEER?

Background: Evaluating the incidence and spatial dynamics of chronic wildlife diseases requires long-term studies that may be difficult to conduct in natural populations due to financial and logistical constraints. Computer simulation models provide a practical way to quantify the spatial and temporal patterns of chronic diseases in wildlife, evaluate alternative transmission mechanisms, predict the spread of the infectious agents across the landscape, and identify viable management options.

Objectives: Jennelle et al. (2014) used hunter harvest data from 2002–2013 to investigate chronic wasting disease dynamics in a Wisconsin white-tailed deer population and explore how alternative management strategies might affect disease spread.

Study Design: Mathematical (simulation) modeling based on cross-sectional data.

Methods: Infection status was determined by immunohistochemistry or ELISA testing of brain stem or retropharyngeal lymphatic tissue collected at deer check stations. Data on 16,773 hunter harvested white-tailed deer from southwestern Wisconsin were obtained from the Wisconsin Department of Natural Resources (WI DNR). A total of 15,136 records, obtained between October and January 2002–2011, were used to develop a deterministic simulation model of CWD transmission and spread under different harvest strategies, and 1637 records obtained between October and January 2011–2013 were used to test model predictions.

Results: CWD was diagnosed in 958 of 16,773 deer, yielding an overall CWD prevalence of 5.7%. Yearly prevalence of CWD infection increased from 3.3%–12.8% over the 11 yearly hunting seasons under study. The computer simulations predicted that in the next decade CWD prevalence would increase to relatively high levels (25% in females and 50% in males) in the

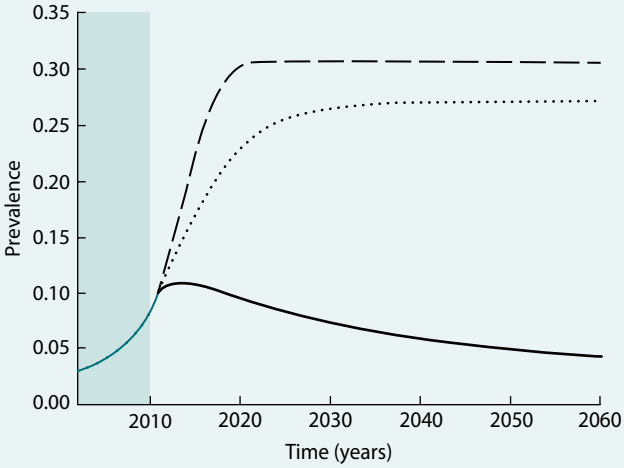

FIGURE 1.1 Predicted CWD prevalence in Wisconsin white-tailed deer populations using transmission estimates from the best-fitting model. Three strategies were considered, including male-focused harvest rates (solid line; female = 25%, male = 50%), herd-control harvest (dotted line; female = 28%, male = 22%), and female-focused harvest (dashed line; female = 50%, male = 25%). Note that the herd-control harvest strategy represents an average of the existing harvest conditions in the south-central core of WI during the 2002–2010 harvest seasons (shaded area). (From Jennelle CS et al. *PLOS ONE* 2014; 9(3): e91043, 1–12. With permission.)

absence of significant management strategies to reduce infection rates. Of the three harvest strategies evaluated, only male-focused harvest succeeded in reducing CWD prevalence below current levels (Figure 1.1).

Conclusions and Significance: Predicted prevalence was reduced because this strategy removed animals from the highest prevalence class (reducing infection rates), while allowing dilution of population-level CWD prevalence by recruitment of more females and stabilizing post-harvest deer density after 20 years. Given increased discovery and distribution of CWD throughout North America, insights from this study will be valuable to management agencies and to the general public concerned about the impacts of CWD on white-tailed deer populations.

FOLLOW-UP QUESTION 1.4

The authors predict that implementation of a male-focused harvest strategy would reduce CWD prevalence to 2.5% by 2110. Surveillance would thus become critical for identifying areas of CWD persistence. Based on these findings, how many deer should be sampled to be 95% and 99% sure of detecting at least one CWD-affected deer if the prevalence were 2.5%? (*Hint:* See Chapter 9, "Statistical Significance for a clue.") See Answer 1.4 at the end of this chapter.

1.4.3 Learn to Read the Medical Literature Critically

Veterinary journals and online electronic resources play an important role in keeping practitioners abreast of current medical knowledge. An international survey of 2137 veterinary clinicians and non-clinicians from 78 countries (Huntley et al., 2016) reported that journals were most commonly used (65.8%, 75.6%), followed by online resources (58.7%, 55.9%). Respondents listed a total of 518 journals and 567 online sources that they read. The most accessed electronic resource reported was the Veterinary Information Network (VIN; 25.6%) for clinicians and PubMed (7.4%) for

non-clinicians. Examples of information sought by clinicians include reports of new and emerging diseases, diagnostic test performance, risk factors for disease and injury, and prognosis with or without medical intervention. The usefulness of this information ultimately depends on the adequacy of the study design and the analysis and interpretation of the data (Buhles and Kass, 2012).

The reader needs to be an astute critic of published studies to determine which therapies are effective and warrant adoption and which do not.

A variety of study designs are used in clinical research (Smith, 1988; Table 1.4). Each has inherent strengths and weaknesses (Table 1.5). Given the effect that chance, bias, and confounding factors can have on the validity of conclusions derived from clinical research, students must learn to evaluate this important resource critically. Further, readers must learn to distinguish between association and causation, and statistical significance versus clinical significance. The effectiveness of veterinary clinical research can be enhanced by choosing epidemiologic study designs appropriate for the clinical issue being examined, and through more rigid adherence to accepted norms for expressing the findings from such studies. The reader needs to be an astute critic of published studies to determine which therapies are effective and warrant adoption and which do not (Buhles and Kass, 2012).

TABLE 1.4

Key for Classification of Study Designs

1. Subjects under study experienced experimentally induced disease, condition, or intervention	Experimental disease
Subjects under study experienced naturally occurring disease, condition, or intervention	Go to 2
2. Fewer than 10 animal units (individuals, herds, etc.) or outbreaks examined	Case report
Ten or more individuals or outbreaks examined	Go to 3
3. Cross-sectional—All observations on a given individual are made at essentially one point in time in the course of that individual's illness	Prevalence survey
Longitudinal—Subjects followed prospectively over a period of time; groups may be formed in the past (from records) or in the present	Go to 4
4. Comparison group absent	Case series
Comparison group present	Go to 5
5. Groups formed based on outcome. Cases selected from an available pool of patients; noncases selected to resemble cases, but not necessarily members of the same population group	Case control study
Groups formed based on exposure	Go to 6
6. No intervention	Cohort study
Intervention	Go to 7
7. Comparison group absent	Uncontrolled clinical trial
Comparison group present	Go to 8
8. Non-random allocation of subjects into treatment and control groups	Non-randomized controlled clinical trial
Random allocation of subjects into treatment and control groups	Randomized controlled clinical trial

TABLE 1.5
Relative Merits of Clinical Research Designs

Study Design	Limitations	Best Application
Case report	Temporal relationships; bias in case selection; statistical validity	Detailed description of uncommon diseases; surveillance
Case series	Temporal relationships; bias in case selection	Frequency of findings in a disease and its clinical course
Prevalence survey	Temporal relationships; measures prevalence, not incidence	Evaluation of diagnostic tests; incrimination of risk or causal factors; outbreak investigation
Case control	Temporal relationships; bias in selection of comparison group	Evaluation of diagnostic tests; incrimination of risk or causal factors; outbreak investigation; rare disease or diseases of long latency
Uncontrolled clinical trial	Time; ethical considerations; no comparison group	Prognosis with treatment
Non-randomized controlled clinical trial	Time; ethical considerations; bias in selection of comparison group	Prognosis with or without treatment; evaluation of new treatments
Randomized controlled clinical trial	Time; ethical considerations	Prognosis with or without treatment; evaluation of new treatments
Experimental disease	Time; availability of animals or other animal models; cost	Proving relationship between risk or causal factors and disease; pathogenic mechanisms

REFERENCES

Bille C, Auvigne V, Bomassi E et al. An evidence-based medicine approach to small animal anaesthetic mortality in a referral practice: The influence of initiating three recommendations on subsequent anesthetic deaths. *Vet Anaesth Analg* 2014;41:249–58. Full text available at: https://www.vaajournal.org/article/S1467-2987(16)30248-3/fulltext

Bille C, Auvigne V, Libermann S et al. Risk of anaesthetic mortality in dogs and cats: An observational cohort study of 3546 cases. *Vet Anaesth Analg* 2012;39:59–68. Full text available at: https://www.vaajournal.org/article/S1467–2987(16)30375-0/fulltext

Buhles W and Kass PH. Understanding and evaluating veterinary clinical research. *J Am Anim Hosp Assoc* 2012;48:285–298.

Fletcher RH, Fletcher SW, and Wagner EH. *Clinical Epidemiology—The Essentials*, 1st ed. Baltimore: Williams and Wilkins; 1982.

Huntley SJ, Dean RS, Massey A et al. International evidence-based medicine survey of the veterinary profession: Information sources used by veterinarians. *PLOS ONE* 2016;11:1–18. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4961404/

Jennelle CS, Henaux V, Wasserberg G et al. Transmission of chronic wasting disease in Wisconsin white-tailed deer: Implications for disease spread and management. *PLOS ONE* 2014;9:1–12. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3962341/

Krone LM, Brown CM, and Lindenmayer, JM. Survey of electronic veterinary medical record adoption and use by independent small animal veterinary medical practices in Massachusetts. *J Am Vet Med Assoc* 2014;245:324–332. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4782149/

McKenzie BA. Is complementary and alternative medicine compatible with evidence-based medicine? *J Am Vet Med Assoc* 2012;241:421–6.

Meléndez-Lazo L, Ordeix M, Planellas J et al. Clinicopathological findings in sick dogs naturally infected with *Leishmania infantum*: Comparison of five different clinical classification systems. *Res Vet Sci* 2018;117:18–27.

Mills DS, Redgate SE, and Landsberg GM. A meta-analysis of studies of treatments for feline urine spraying. *PLOS ONE* 2011;6:1–10. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3078130/?tool=pubmed

Morens D. When was epidemiology born? *The Epidemiology Monitor* 2003;20:3,6,9.

Mosby's Medical Dictionary, 2009; Mosby, Inc. St Louis, Missouri.

NLM, Medical Subject Heading (MeSH), U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894; 2018. Available online at: https://meshb.nlm.nih.gov/search
Ramey D. "Alternative" therapies: Doing the right thing. *Vet Forum* 2003a;20:30–31.
Ramey DW. Regulatory aspects of complementary and alternative veterinary medicine. *J Am Vet Med Assoc* 2003b;222:1679–1682.
Robinson NJ, Brennan ML, Cobb M et al. Agreement between veterinary patient data collected from different sources. *Vet Journal* 2015;205:104–106. Full text available at: https://www.sciencedirect.com/science/article/pii/S1090023315001653?via%3Dihub
Rollin BE. The use and abuse of Aesculapian authority in veterinary medicine. *J Am Vet Med Assoc* 2002;220:1144–1149.
Sackett DL, Haynes RB, Guyatt GH et al. *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd ed. Boston: Little, Brown and Co; 1991.
Sackett DL, Richardson WS, Rosenberg W et al. *Evidence-based Medicine. How to Practice and Teach EBM*. New York: Churchill Livingston; 1997.
Schwabe CW, Riemann HP, and Franti CE. *Epidemiology in Veterinary Practice*. Philadelphia: Lea and Febiger; 1977.
Smith, RD. Veterinary clinical research: A survey of study designs and clinical issues appearing in a practice journal. *J Vet Med Educ* 1988;15:2–7.
The American Heritage Medical Dictionary. 2007; Houghton Mifflin Harcourt; Boston, Massachusetts.
Wikipedia contributors. Epidemiology. Wikipedia, The Free Encyclopedia. April 2, 2019. Full text available at: https://en.wikipedia.org/wiki/Epidemiology
World Health Organization. *Epidemiology*; 2019, available at: https://www.who.int/topics/epidemiology/en/

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 1.1: The numerical value of the odds ratio (OR, a measure of risk) conveys the magnitude of the risk, but not its statistical significance. In order for an OR to be statistically significant, the *p* value (likelihood of obtaining the observed OR by chance) should be less than 0.05, and the 95% confidence interval (range over which the OR is likely to occur) should not include 1.00 (no difference). The odds ratios for **three variables** (ASA categories 4 and 5, and geriatric patients) in Table 1.2 meet these criteria.

Answer 1.2: As this was a meta-analysis of published research results, the investigators had no control over the study designs. However, they did exclude some studies that failed to meet certain criteria to minimize the effect of bias. Among the factors that could be better controlled for in a randomized controlled clinical trial of treatments for feline urine spraying are: standard case definition, owner compliance with treatment regimens, blinding of owners and investigators with regard to treatment status, placebo effect, and the definition of treatment success.

Answer 1.3: When interpreting the findings from a diagnostic workup, the absence of a finding that is usually present in individuals afflicted with a particular disease (a sensitive finding) can be used to rule out the disease in question, or at least place it lower on the differential list. In the case of leishmaniasis, the investigators found that cutaneous lesions were present in 78.4% of patients. Thus, **absence of cutaneous lesions** would suggest that a patient does not have leishmaniasis. Note that the presence of skin lesions cannot be used to rule in a diagnosis of leishmaniasis without knowing how many other diseases on the differential list might have similar lesions.

Answer 1.4: One way to look at this problem is to ask what the chance is of getting only negative test results (referred to as an "extreme outcome") when testing a series of animals in a herd. If the prevalence of disease is 2.5%, then the chance of a negative test result on a randomly selected animal is 0.975. For two randomly selected animals, there is a 0.975^2 chance that both tests are negative, or 0.951. Repeatedly raising the power of the equation (number of animals tested) eventually yields probabilities of missing the disease of 0.05 and 0.01, and probabilities of detecting it in at least one of the animals of 0.95 and 0.99, or 95% and 99%. The corresponding numbers of animals that need to be tested are **119,182**. The actual equation used to arrive at these estimates can be found in Chapter 9.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

2 Defining the Limits of Normality

2.1 INTRODUCTION

> ***Personally, I have always felt that the best doctor in the world is the veterinarian. He can't ask his patients what is the matter…he's just got to know.***
>
> **—Will Rogers**
> **(Pediatricians would probably take issue with this.)**

Although the way in which we gather data may at times differ, the process of veterinary and human medical decision-making is basically the same and consists of at least four steps. First, **subjective data** are collected, such as alertness, attitude, evidence of pain, etc. These data are based on our own observations and those of the owner. **Objective data** are also collected; indices include temperature, pulse, respiration, results of parasitological examinations, complete blood counts, radiographs, etc. These data are then interpreted as either normal ("within normal limits," "unremarkable," "noncontributory") or abnormal in light of our past experience and the medical literature, and we arrive at an **assessment** (or, in some cases, "appreciation") of the problem. Depending on this assessment, we then devise a **plan** that may be a more complete workup, a rule-out of other possible diagnoses, a treatment, or client education (Seibert and Walker Smith, 2005).

> ***Although the way that we gather data may at times differ, the process of veterinary and human medical decision-making is basically the same and consists of at least four steps.***

At this point the astute reader will have realized that the acronym for this process (subjective data, objective data, assessment, and plan) is SOAP. SOAPs are part of the **problem-oriented medical records** system that provides a formal way of recording subjective and objective data about a patient. From these databases, patient problems are isolated and defined. All recognized problems, past and present, are assessed and listed as a "problem list," and plans for the management of each problem are then recorded.

In this chapter we first review the properties of clinical measurements and their distributions within animal populations. Next, we develop criteria by which abnormal values for clinical measurements are recognized, including normal reference ranges.

2.2 PROPERTIES OF CLINICAL MEASUREMENTS

Practitioners are continually collecting, categorizing, and quantifying biological data about their patients. In the hospital environment, these data are categorized as patient history, clinical signs, and screening/definitive tests. The important point to remember is that clinical data alone mean nothing until interpreted in the context of expected values for the population. Clinical assessment is

based on the degree to which patient data differ from population "norms" and match expectations for particular disease syndromes. The response to the treatment plan is assessed by the rate and degree to which clinical findings return to normal population values. In this section we examine the factors that influence the confidence we place in clinical measurements.

2.2.1 Signs and Symptoms: Objective versus Subjective Data

The following are definitions from *Mosby's Medical Dictionary* (2009):

- A **sign** is "an objective finding as perceived by an examiner."
- A **symptom** is "a subjective indication of a disease or a change in condition as perceived by the patient."

> ***Clinical data alone mean nothing until interpreted in the context of expected values for the population.***

It has been argued that because our patients cannot talk, veterinarians rely only on signs to assess the clinical condition and progress of patients. Animals are generally more stoic than humans and may not exhibit behavioral alterations until the condition has progressed quite far. Yet, our assessment of a patient's health may include subjective evidence that fits the definition of symptoms. Furthermore, we often use the terms *symptomatic* or *asymptomatic* to describe the presence or absence of evidence of disease.

It is important to recognize subjective data as subjective and ensure that measures have been taken to reduce the influence of personal bias in clinical measurements. Behavioral characteristics are an example of subjective data used to describe animals.

EXAMPLE 2.1: ARE THERE BEHAVIORAL DIFFERENCES BETWEEN DOGS OBTAINED AS PUPPIES FROM PET STORES ("PUPPY MILLS") VERSUS NONCOMMERCIAL BREEDERS?

Background: Most puppies sold by pet stores in the United States are purchased from brokers, who may themselves be breeders but overwhelmingly acquire their puppies from high-volume breeding facilities, or commercial breeding establishments (CBEs), located throughout the United States. Conditions in the CBEs, which supply tens of thousands of puppies to retail pet stores each year, vary widely. Conditions in CBEs range from modern, clean, and well-kept to squalid, noxious, and gravely detrimental to animal health and welfare.

Objectives: McMillan et al. (2013) tested the hypothesis that dogs obtained as puppies from pet stores would have an increased prevalence of behavioral problems, compared with dogs obtained as puppies from noncommercial breeders (NCBs).

Study Design: Cross-sectional survey of pet owners.

Methods: Owner-reported prevalence of behavioral characteristics in dogs obtained as puppies from pet stores was compared with that of dogs obtained as puppies from noncommercial breeders. A convenience sample (participants were self-selected) of pet dog owners who reported obtaining puppies from either breeders (n = 5,657) or pet stores (413) completed the online version of the Canine Behavioral Assessment and Research Questionnaire (C-BARQ), which uses ordinal scales (see discussion of scales below) to rate either the intensity or frequency of their dog's behavior. Hierarchic linear and logistic regression models were used to

analyze the effects of source of acquisition on behavioral outcomes when various confounding and intervening variables were controlled for.

Results: Pet store–derived dogs received significantly less favorable scores than did breeder-obtained dogs on 12 of 14 behavioral variables measured; pet store dogs did not score more favorably than breeder dogs in any behavioral category. Compared with dogs obtained as puppies from noncommercial breeders, dogs obtained as puppies from pet stores had significantly greater aggression toward human family members, unfamiliar people, and other dogs; greater fear of other dogs and nonsocial stimuli; and greater separation-related problems and house soiling.

Conclusions and Significance: Obtaining dogs from pet stores versus noncommercial breeders represented a significant risk factor for the development of a wide range of undesirable behavioral characteristics. Until the causes of the unfavorable differences detected in this group of dogs can be specifically identified and remedied, the authors could not recommend that puppies be obtained from pet stores.

FOLLOW-UP QUESTION 2.1

What are the possible sources of bias in this report? (*Hint*: See Chapter 6, "Risk Assessment and Prevention" and Chapter 12, "Establishing Cause for clues.") See Answer 2.1 at the end of this chapter.

2.2.2 Scales

Clinical data are of three types: nominal, ordinal, or interval. **Nominal data** can be placed into discrete categories that have no inherent order. Another name for nominal data is **categorical data**. Clinical phenomena that fall into this category are either inherent characteristics of an animal (e.g., name, species, breed, sex, and coat color) or are discrete events (e.g., fracture, birth, death).

Clinical data are of three types: nominal, ordinal, or interval.

Ordinal data are categorical data with an obvious order that can be ranked, but the intervals are not uniform in size. Examples are degrees of aggression, depression, pain or anxiety, degrees of dehydration or incoordination, and severity of respiratory sounds or cardiac murmurs. One student wrote in a canine patient's progress report: "On an alertness scale of 1 to 5, give him a 3."

Data that are ordered and for which the sizes of the intervals are equal are called **interval**. Another name of interval data is **continuous data**. Examples are weight, rectal temperature, packed cell volume, and leukocyte count. The size of the intervals depends on the precision of instruments used to make the measurements.

Most interval level scales used in medicine have absolute (mathematically meaningful) zero points; e.g., a value of zero means absolute zero of the quantity being measured. Examples are body weight, heart and respiratory rates, blood chemistries and differentials, etc. A negative value is not possible. In some cases, such as the Fahrenheit or Celsius temperature scales, zero is simply an arbitrary point whose value happens to be called "zero." On these scales, zero does not represent an absolute absence of the factor being measured, and negative values are possible. Interval scales that have absolute zero points are sometimes referred to as ***ratio*** scales. Ratio scales permit the meaningful calculation of ratios. For example, if an animal's packed cell volume (PCV) increases from 10% to 20% in response to treatment, it is legitimate and meaningful to say that the PCV has doubled. If, on the other hand, the high temperatures on two successive days are 4°C

and 8°C, it makes no sense to conclude that the second day is twice as warm as the first because the zero point from which 4°C and 8°C are starting out is only an arbitrary marker on a scale that potentially extends all the way down to about −273°C. In order to make such ratio judgments concerning temperatures, we would have to use a scale, such as the Kelvin scale, whose zero point does mark an absolute zero level of temperature. Since ratio-level variables are treated the same as interval-level variables for all other statistical purposes, they will be considered interval data throughout this text.

It is not uncommon for interval-level data to be reduced to the ordinal or categorical level in clinical records. For example, a hospital admission record may divide age and body weight into unequal interval classes (age: 0–2 weeks, 2 weeks–2 months, 2–6 months, etc.; weight: 0–1 lb, 1–5 lb, 5–15 lb, etc.). These lower scales of measurement precision can be convenient for summarizing large amounts of information into clinically meaningful categories. However, useful information may be lost in the process. For example, a follow-up study of the prognostic values of animal age or weight for a specific condition may be impossible without precise interval-level data. Therefore, if time and other limitations permit, information should be recorded at the same level as it was measured.

Sometimes ordinal data are recorded on an interval-level scale and then analyzed statistically as if they were truly interval. This is an inappropriate conversion of data, as it may misrepresent the magnitude of differences among individual measurements. Furthermore, the raw data are often subjective in nature and do not meet the measurement and reproducibility criteria of interval-level data.

EXAMPLE 2.2: HOW EFFECTIVE ARE TREATMENTS FOR CANINE CONSPECIFIC COPROPHAGY?

Background: Conspecific coprophagy, the ingestion of an animal's own feces or that of another of the same species, is a problem behavior in dogs that is both unattractive and unhygienic.

Objectives: Wells (2003) conducted a clinical trial to assess the effectiveness of two treatments on the reported severity of auto coprophagy in a group of dogs referred for behavior therapy.

Study Design: Randomized controlled clinical trial.

Methods: Twenty-eight client-owned labrador retriever dogs with a history of coprophagy were included in the study. A single breed was included to reduce the possibility of breed differences influencing the results. The dogs were divided at random into two groups. Half were treated with a citronella spray collar that emits a cloud of spray together with an audible hiss under the dog's nose whenever the collar is triggered remotely by the owner. The remainder were exposed to sound therapy via a hand-held alarm that emits a 115 dB screech whenever the trigger of the propellant spray is pressed, thereby interrupting the dog's behavior. To assess the relative efficacy of the treatments, owners rated the severity of their dog's feces eating for a week before the study began, during each of 3 weeks of treatment, and at the end of a fourth week during which they had not been treated. Severity of coprophagy was scored on a scale ranging from 1 to 4, where 1 indicated 0%–25% of available feces were eaten, 2 indicated 26%–50% of available feces were eaten, 3 indicated 51 to 75% of available feces were eaten, and 4 indicated 76%–100% of available feces were eaten, with a concomitant loss of precision.

Results: The owners of the dogs treated with the spray collar reported a lower mean severity of feces eating in their pets (1.35 ± 0.15 s.e.) than the owners of the dogs treated by sound therapy (2.98 ± 0.18 s.e.).

Conclusions and Significance: The author concluded that a citronella spray collar may be a more effective means of reducing the severity of feces eating in domestic dogs than sound

therapy. From an animal welfare point of view, it may be wiser to use a spray collar that emits an odorless, rather than a citronella-scented, spray because there is evidence that both types of collar are equally effective in disrupting unwanted behavior.

FOLLOW-UP QUESTION 2.2

This scoring system is a clear example of converting broad ordinal data (0% to 25%, 26% to 50%, etc.) to a fixed interval (1, 2, etc.). Can you think of scenarios where conversion of the ordinal data in this study to interval-level data could bias the interpretation of results? (*Hint*: See Chapter 9, "Chance for a clue.") See answer 2.2 at the end of this chapter.

TABLE 2.1
Clinical Assessment of Anemia in the Dog and Cat

Nominal	Breed, sex, diet, history of drug administration or recent infection, existence of a heart murmur or hemorrhages
Ordinal	Color of mucous membranes, grade of heart murmur
Interval	Age, cardiac and respiratory rates, packed cell volume, complete blood count, frequency distribution of erythrocyte morphologic types, total plasma protein

Source: Straus JH. Anemia. In, Fenner WR (ed). *Quick Reference to Veterinary Medicine*. Philadelphia: J B Lippincott Co; 1982.

The differences among nominal, ordinal, and interval-level variables can be appreciated in Table 2.1, which summarizes the clinical assessment of canine and feline anemia.

2.2.3 Clinical Staging

Clinical staging is another expression of the degree of abnormality. Separation of patients based on the severity of their condition is necessary before comparing such things as diagnostic tests, prognosis, and response to treatment.

One internationally recognized form of clinical staging is the TNM Classification of Tumors in Domestic Animals (Owen, 1980), which was established by an international consultation sponsored by the World Health Organization (WHO) Programme on Comparative Oncology. The staging criteria were modeled after a classification system established in 1968 for tumors in humans. The principal purpose of international agreement on clinical staging of animal tumors is to provide a method of conveying clinical observations without ambiguity. The system arose from the fact that survival rates were higher for localized, compared with disseminated, tumors. Before establishment of the TNM staging system, these groups were often referred to as "early cases" and "late cases," implying some regular progression with time.

EXAMPLE 2.3: WHAT PROGNOSTIC CRITERIA ARE BEING USED BY ONCOLOGISTS FOR CLINICAL STAGING OF DOGS WITH LYMPHOMA?

Background: Clinical substage is frequently reported to be of prognostic value in dogs with lymphoma, yet formal criteria for defining this parameter are lacking. The World Health Organization TNM Classification of Tumors of Domestic Animals simply defines substage as the absence or presence of systemic signs (substages a and b, respectively).

Objectives: Barber and Weishaar (2016) designed a survey to query veterinary oncologists on the criteria they use to determine clinical substage in practice.

Study Design: Cross-sectional survey.

Methods: Questionnaires were mailed to 212 diplomates of the American College of Veterinary Internal Medicine in the subspecialty of oncology and residents eligible for the oncology

TABLE 2.2
The Proportion of Respondents That Identified Each Clinical Factor as a Criterion for Assigning Dogs with Multicentric Lymphoma to Clinical Substage b

Clinical Parameter	Number of Responses	Positive Responses (%)
	Gastrointestinal	
Inappetence	122	98
Vomiting	122	100
Diarrhea	122	97
Constipation	121	23
	Constitutional	
Attitude	122	95
Lethargy	121	89
Weakness	122	94
	Nutritional	
Weight loss	120	88
Body condition	118	73
Muscle wasting	97	56
	Respiratory	
Dyspnea	121	90
Coughing	120	64
Sneezing	122	20
Snoring	122	15
Nasal discharge	121	43
	Neurologic	
Seizures	121	86
	Metabolic	
Polydipsia	121	58
Polyuria	121	59
Hypercalcemia	121	78
Fever	122	78
	Other	
Infections	121	59
Bruising	122	70
Ocular changes	121	74
Excessive shedding	121	7
Owner's assessment	122	95

Source: Barber LG, Weishaar KM. *Vet Comp Oncol* 2016;14:32–39. With permission.

subspecialty examination. They were queried on the criteria they used to determine clinical substage of dogs with multicentric lymphoma. The affirmative responses for each clinical attribute were tabulated as a percentage of the total number of responses for that parameter.

Results: Questionnaires were returned by 124 recipients (response rate 58%), of which 122 were evaluable. Criteria for clinical staging by respondents varied considerably, depending in part on the subjectivity of the factors being considered. The proportion of respondents that identified each clinical factor as a criterion for clinical substage b (presence of clinical signs) designation is summarized in Table 2.2. Gastrointestinal, constitutional, and respiratory signs were the most commonly identified clinical factors, with greater than 90% of respondents indicating that inappetence, vomiting, diarrhea, changes in attitude, weakness, and dyspnea were integral in assigning clinical substage. Nevertheless, more than three-quarters of respondents also considered metabolic, neurologic, and nutritional parameters when making this determination. For most factors, respondents reported that mild-to-moderate severity of clinical signs was sufficient for substage b designation.

Conclusions and Significance: This survey should be considered an exploratory investigation that can inform future studies. Prospective inquiries may identify specific clinical signs and their severity that are most closely correlated with prognosis—either as independent predictors of outcome or as markers of disease factors indicative of aggressive biologic behavior or poor response to therapy. In addition, systematic evaluation of selected parameters may assist in developing algorithms for assessing interrelated clinical signs.

FOLLOW-UP QUESTION 2.3

Which of the clinical parameters in Table 2.2 are interval-level variables? See answer 2.3 at the end of this chapter.

Clinical staging is necessary, but definitions are only as good as the criteria used to construct them. Furthermore, clinical staging is based on the present state of knowledge, and most systems will require modification in the future.

2.2.4 Validity and Reliability

Validity and reliability are terms that have been used to describe the quality of clinical measurements. **Validity** (or accuracy) describes the degree to which a measurement reflects the true status of what is being measured. **Reliability** is a measure of the repeatability or reproducibility of a clinical measurement. Reliability is sometimes referred to as **precision**.

Validity and reliability are relatively easy to establish when measurements can be compared with some accepted standard. Examples are blood chemistry measurements in which instruments are calibrated with known standards. Another example may be serodiagnostic tests in which subsequent culture or necropsy may confirm the presence or absence of disease. Validity and reliability are more difficult to establish for other clinical measurements that rely on our senses and for which no physical standards exist. Examples might be the validity of our estimate of the severity of pneumonia based on auscultatory findings, or the reproducibility of pain scores assigned to patients by different observers, e.g., **interobserver variability**.

Validity may be independent of reliability. Repeated serologic tests on a serum sample, for example, may give consistently valid (accurate) results, but titers may vary considerably about the true value. In contrast, an improperly functioning thermometer can be very reliable, but systematically off the mark (inaccurate).

The **coefficient of variation** (CV) is frequently used to express the precision of clinical measurements. The CV is equal to the standard deviation of a set of measurements divided by their mean and is usually expressed as a percentage. The CV therefore represents the percentage variation of a set of measurements around their mean and provides a useful index for comparing the precision of different instruments, individuals, or laboratories.

2.2.5 Variation

There are two major sources of variation in clinical measurements. One is associated with the act of measurement itself, while the other is associated with biological variation within and among individuals. Clinicians should be aware of potential sources of variation to avoid erroneous conclusions about data in a given situation.

2.2.5.1 Measurement Variation

Measurement variation may be due to variation in the way samples are handled and processed, the performance of the instruments being used, or the observers themselves. It can be thought of as the variation recorded during repeated measurements of the same parameter in an individual, irrespective of other members of the population.

EXAMPLE 2.4: HOW RELIABLE ARE TESTS FOR DETERMINING ADEQUACY OF TEAR PRODUCTION IN DOMESTIC CATS?

Background: A variety of tests are available for evaluating the adequacy of tear production in animals that have signs of keratoconjunctivitis. Despite this, reference ranges for many of the tests commonly used to assess the human and canine tear film have not been established for healthy cats. As a result, it is likely that clinically important alterations in tear film health may be undiagnosed or underdiagnosed in cats.

Objectives: Sebbag et al. (2015) conducted a study to determine reference values, intertest correlations, and test-retest repeatability of Schirmer tear test 1 (STT-1), phenol red thread test (PRTT), tear film breakup time (TFBUT), tear osmolarity, and meibometry in healthy cats.

Study Design: Cross-sectional (n = 120) and case series (n = 40).

Methods: Each test was performed once in 120 healthy domestic cats. The Pearson correlation was used to assess correlation among tests. To evaluate test-retest repeatability, each test was repeated by the same evaluator under identical conditions 1 week later on 40 of the cats (80 eyes) chosen at random (Figure 2.1). Intraclass correlation coefficients (ICCs) and 95% limits of agreement (LOA) were used to evaluate test-retest repeatability.

Results: Generally poor correlation among tests suggested that thorough tear film analysis requires performance of multiple tests in concert. The STT-1 and PRTT values were positively correlated. Age was weakly associated with TFBUT and osmolarity. Meibometry measurements (meibometry units; MU) were higher for strips that contacted the tear film (285 MU) than for those that touched the eyelid margin only (32 MU). All ICCs were <0.75, and 95% LOA were wide.

Conclusions and Significance: Tear deficiency should be suspected in cats with STT-1 <9 mm/min, PRTT <15 mm/15 s, or TFBUT <9 to 10 seconds. Relatively poor test-retest repeatability should be considered when repeated tests are used to monitor tear film dysfunction and response to treatment, and suggests that thorough tear film analysis requires performance of multiple tests in concert. The variability in test results could influence the clinical diagnosis

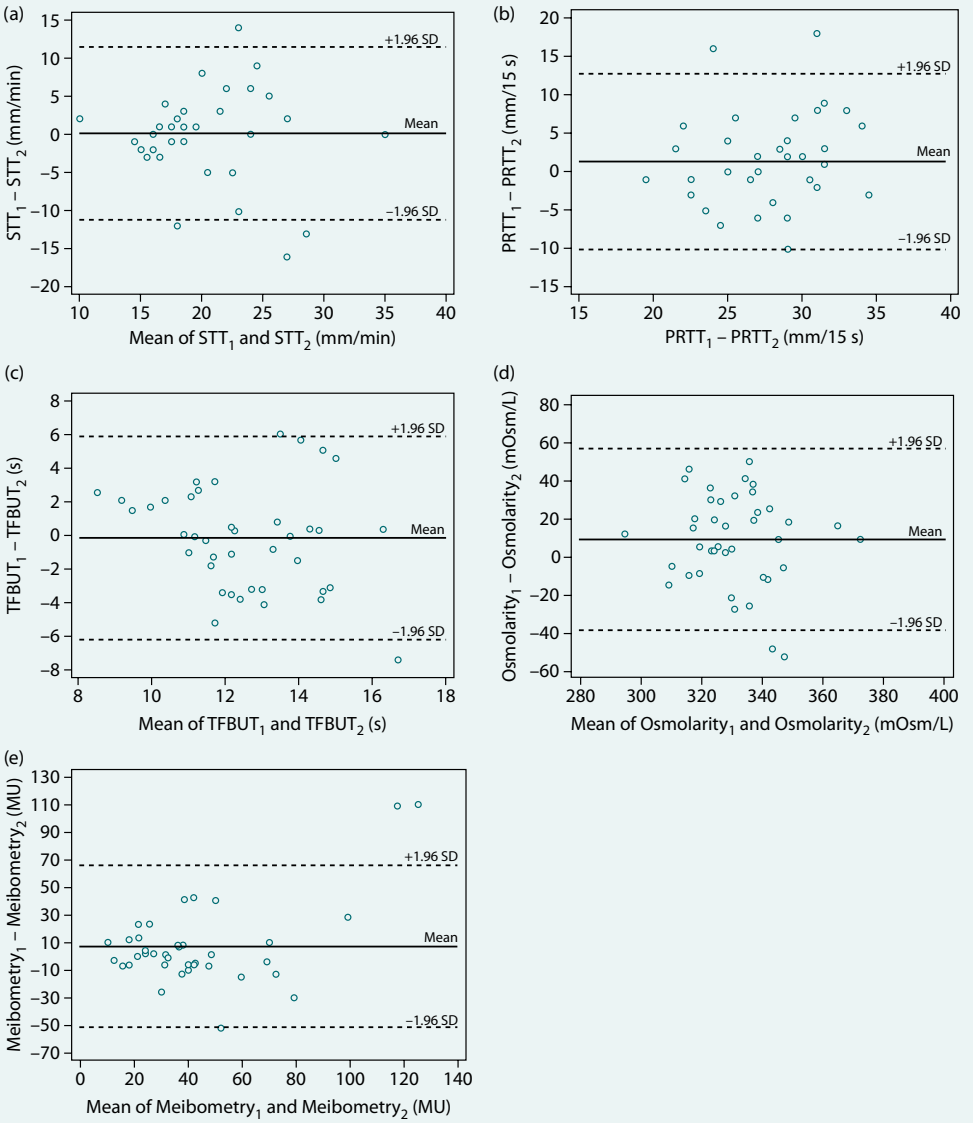

FIGURE 2.1 Bland-Altman plots displaying test-retest repeatability for the Schirmer tear test (a), phenol red thread test (b), tear film breakup time (c), tear osmolarity (d), and peak meibometry (e) assays in 40 healthy cats. The vertical axis represents the difference between repeated measurements, and the horizontal axis plots the mean value for the two sessions. The mean of the differences is represented as the solid horizontal line intersecting the vertical axis and should be close to zero. The dashed lines represent the 95% limits of agreement. (From Sebbag L et al. *J Am Vet Med Assoc* 2015;246:426–435. With permission.)

of keratoconjunctivitis sicca and the subsequent interpretation of response to treatment, as well as the interpretation and repeatability of research data.

FOLLOW-UP QUESTION 2.4

Based on the findings depicted in Figure 2.1, which of the five tests for tear production exhibits the highest repeatability? See answer 2.4 at the end of this chapter.

2.2.5.2 Biological Variation

Biological variation is defined as the random fluctuation of a test result around a central point, and is composed of the within-subject variation, reflecting changes occurring in the same individual over time, and between-subject variation, representing the difference between individuals (Falkenö et al., 2016; Campora et al., 2018). Biological variation can manifest at all levels of an animal population. The histopathologic description of a biopsy, for example, may vary depending on the region of the affected lesion or the organ from which it is taken. Clinical measurements vary over time within an individual. In some cases this variation may be cyclic, such as hormone levels, heartworm microfilaria counts, or body temperature. In others it varies with each patient. The utility of population-based reference intervals (reference ranges), interpretation of serial test results, and establishment of analytical quality specifications are all influenced by biological variation.

The utility of population-based reference intervals can be assessed by comparing the within-subject coefficient of variation (CV_I) to the between-subject coefficient of variation (CV_G) and calculating an analyte's **index of individuality** (II). The analytical coefficient of variation (CV_A) must also be considered, as it reflects analytical precision and contributes to the overall variation. When this "analytical noise" is excessive, it may mask physiological variations, thus interfering with patient data interpretation. CV_I and CV_G for many species and measurands can be obtained from the literature or online databases. Individual laboratories can estimate their CV_A from replicate analyses or from quality control studies using either pooled patient samples or quality control materials, with concentrations at or near clinical decision limits of interest (Campora et al., 2018).

II may be calculated using either of the following formulae:

$$II = \frac{CV_G}{\sqrt{CV_I^2 + CV_A^2}} \quad (1)$$

$$II = \frac{\sqrt{CV_I^2 + CV_A^2}}{CV_G} \quad (2)$$

As the two formulae yield reciprocal numerical results, it is imperative that investigators report the formula used to determine II (Freeman et al., 2017). Formula (1) is preferred, as the results are more intuitive, e.g., higher II = higher individuality, and when used, the following interpretation applies (Campora et al., 2018; Freeman et al., 2017):

- *II > 1.67*: High degree of individuality. The distribution of values from a single individual will cover much of the entire distribution of the population-based reference interval. When individuals have values that are even slightly unusual results for them, there is a high probability that these values will fall outside the reference interval. Consequently, the use of population-based reference intervals is of low utility. Many routine clinical laboratory measurements show high individuality (see Example 2.5 below).
- *II < 0.7*: Low levels of individuality. The dispersion of values for any individual will span only a small part of the population-based reference interval. Individuals may have values that are very unusual for them but still lie within the reference interval. The use of a traditional population-based reference interval is generally useful.
- *II between 0.7 and 1.67*: Moderate individuality. The use of both population- and patient-based reference intervals may be helpful in diagnosis.

When serial test results are being used to monitor disease progression or treatment effects, the use of patient-based reference intervals is appropriate.

Veterinary medicine is unique in that practitioners deal with disease at both the individual and herd levels. Although the effects of biologic variation on herd data can be moderated by taking larger

sample sizes, there is little the practitioner can do to reduce the effects of biological variation when interpreting tests on individual patients. As a rule, rigid adherence to test protocols is the single most important way to reduce overall test variation.

EXAMPLE 2.5: HOW USEFUL ARE POPULATION-BASED REFERENCE INTERVALS FOR DIAGNOSTIC TESTING IN CATS?

Background: Biological variation is defined as the random fluctuation of an analyte around a homeostatic setting point, and it is composed of within-subject variation (CV_I), reflecting changes occurring in the same individual over time, and between-subject variation (CV_G), representing the difference between individuals. The applications of data on biological variation include setting of analytical quality specifications, assessment of the utility of population-based reference intervals, and evaluation of the significance of change in serial results.

Objectives: Falkenö et al. (2016) investigated the utility of population-based reference intervals in cats by comparing the degree of within-subject variation to the degree of between-subject variation and calculating an analyte's index of individuality (IND_I).

Study Design: Case report (fewer than 10 individuals examined).

Methods: The authors investigated the biological variation of 19 biochemistry analytes and thyroxine (total T4), measured in serum from seven clinically healthy domestic cats sampled once weekly for 5 weeks. Samples were frozen and analyzed in random order in the same analytical run. The generated variance components for each analyte were converted into their corresponding analytical precisions (CV_A), within-subject variations (CV_I), and between-subject variations (CV_G), using the overall mean value. The utility of population-based reference intervals was assessed by comparing the degree of within-subject variation to the degree of between-subject variation and calculating an analyte's index of individuality (IND_I) using the formula $IND_I = [CV_A^2 + CV_I^2]^{1/2}/CV_G$. **In this case, (formula 2 above) low IND_I indicates limited use of population-based reference intervals.**

Results: The smallest CV_I and CV_G were found for calcium, chloride, and sodium, whereas the largest values were calculated for bile acids. Zero values for CV_I of two of the analytes (chloride and sodium) precluded estimation of their IND_I values. Of the remaining 18 analytes, 9 (albumin, alkaline phosphatase, alanine aminotransferase, aspartate aminotransferase, cholesterol, creatinine, phosphate [phosphorus], total protein, total T4) had an IND_I of ≤ 0.6, indicating limited use of population-based reference intervals. The index of individuality was ≥ 1.4 for 5 analytes (bile acids, calcium, fructosamine, glucose, potassium), indicating suitability of population-based reference intervals, whereas 4 analytes (CK, iron, magnesium, urea) showed intermediate indices of individuality, suggesting that population-based reference intervals should be used with caution.

Conclusions and Significance: For 9 of 18 analytes measured in this study, population-based reference intervals were determined to be of limited use.

FOLLOW-UP QUESTION 2.5

How might the utility of the population-based reference intervals described above be improved? See answer 2.5 at the end of this chapter.

2.2.5.3 How to Reduce the Effects of Variation

In an effort to reduce variation, it may be useful to distinguish **random variation** from **systematic variation**, or **bias**. Random variation results from the chance distribution of measurements, such as

leukocyte counts in different microscope fields, around an "average" value and will not significantly alter our interpretation of the true status of what is being measured. Inaccuracy due to random variation can be reduced by taking a larger sample size. On the other hand, systematic variation, such as leukocyte counts reported by different technicians, may consistently be biased. In these cases, use of a correction factor may be indicated. This is what we are actually doing when we "blank" an instrument, such as a spectrophotometer, or when we adjust the scale of a chart recorder. As long as these corrections are made carefully and systematically, the validity of the data is not compromised.

Reference ranges for clinical measurements should be determined and expressed by age intervals for each species. For example, plasma protein values are very low in dogs at birth, elevate to the levels seen in the dam after the puppies have nursed, gradually drop during the second 6 months of life, then begin to elevate again after the first year. Maximum levels for this parameter in dogs are reached at about 7 to 10 years, after which the animal will have gradually decreasing values. Leukocyte differential counts in cattle are similar to those in dogs and cats from birth to weaning. After that, they change drastically in the bovine and lymphocytes become the predominant peripheral leukocyte.

2.3 DISTRIBUTIONS

The adage that a picture is worth a thousand words (or numbers) is nowhere truer than in the expression of population data. Data that can be measured on an interval scale, whether continuous or discrete, can be expressed as a "**frequency distribution**." The frequency distribution may be presented as a table or as a graph, referred to as a "**histogram"** or "**frequency polygon**." Frequency distributions may take many forms, but all include at least one scale representing the range of possible values in a distribution, usually divided into intervals, and a second scale depicting the number or proportion of the population that falls within each interval.

A typical histogram is depicted in Figure 2.2, which depicts the distribution of 102 normal canine body temperatures. The size of each interval on the abscissa (x-axis) is 0.2°F. Any other interval could have been chosen as long as it was not smaller than that used to actually record the measurements. The scale on the ordinate (y-axis) depicts the proportion of dogs in each interval.

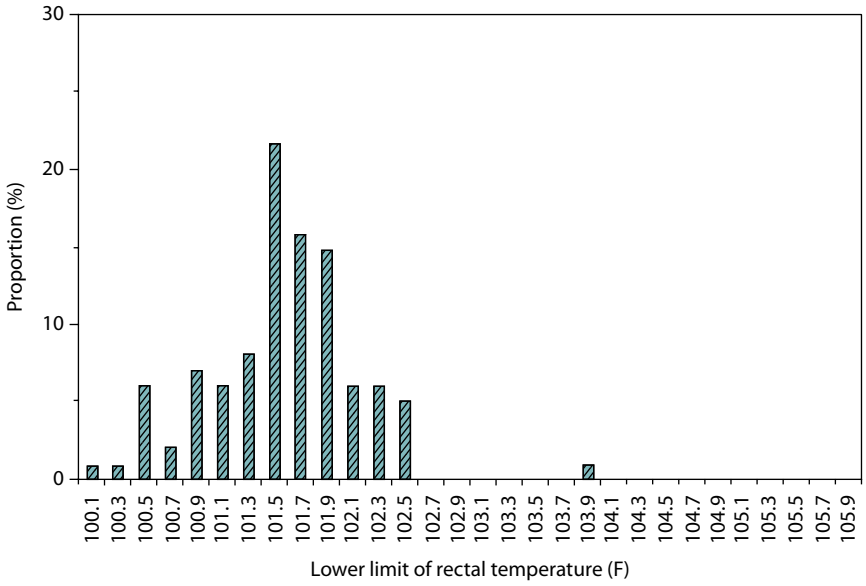

FIGURE 2.2 Frequency distribution of rectal temperatures in 102 normal dogs.

Note that x-axis intervals are retained in the histogram even for temperature values in which the count is zero. Histograms should not be confused with **bar charts**, which differ from histograms in that intervals for values in which the count is zero are omitted. Bar charts are useful for displaying the counts for values of both ordered and non-ordered variables.

2.3.1 Basic Properties of Distributions

Figure 2.2 provides a useful summary of the 102 temperature readings, but it may be helpful to further describe these data, particularly if we wish to compare it with other temperature distributions. Two basic properties of distributions can be used to summarize this data: ***central tendency***, or the middle of the distribution, and ***dispersion***, an index of the spread of the data. There are various ways of expressing central tendency and dispersion. These are summarized in Table 2.3 along with their advantages and disadvantages.

> ***Two basic properties of distributions can be used to summarize data: central tendency and dispersion.***

TABLE 2.3
Expressions of Central Tendency and Dispersion for Frequency Distributions

Expression	Definition	Advantages	Disadvantages
	Measures of Central Tendency		
Mean	Sum of values for observations ÷ number of observations	Well-suited for mathematical manipulation	Easily influenced by extreme values
Median	The point where the number of observations above equals the number below	Not easily influenced by extreme values	Not well-suited for mathematical manipulation
Mode	Most frequently occurring value	Simplicity of meaning; the only way to describe the center of ordinal data	Sometimes there are no, or many, "most frequent" values
	Measures of Dispersion		
Range	From lowest to highest value in a distribution	Includes all values	Greatly affected by extreme values
Standard deviation	The absolute value of the average difference of individual values from the mean	Well-suited for mathematical manipulation	For non-Gaussian distributions, does not describe a known proportion of the observations
Percentile, decile, quartile, etc.	The proportion of all observations falling between specified values	Describes the known proportion of observations without assumptions about the shape of the distribution	Not well-suited for statistical manipulation

Source: Fletcher RH et al. *Clinical Epidemiology - The Essentials*, 3rd ed. Baltimore: Williams and Wilkins; 1996. With permission.

2.3.2 Shapes of Naturally Occurring Distributions

2.3.2.1 Unimodal, Bimodal, and Multimodal

The frequency distribution for a variable can have one or more measurement values with the maximum frequency, or **mode**. The shape of a distribution can be characterized in part by the number of modes it has. A distribution with only one modal value is unimodal, with two modal values is bimodal, etc. In general, a distribution with more than one mode is called multimodal.

2.3.2.2 Symmetry, Skewness, and Kurtosis

Another characteristic of the shape of a distribution is **symmetry** (or its converse, skewness). These properties are reflected in the relationship between the mean, median, and mode of a distribution. In symmetrical distributions, the mean, median, and mode are equal. In positively skewed distributions, the mean is greater than the median, due to extreme values at the upper end of the distribution (often referred to as "skewed to the right"). In negatively skewed distributions, the mean is less than the median, due to extreme values at the lower end of the distribution ("skewed to the left"). **Kurtosis** describes the "peakedness" of a data distribution, e.g., whether the shape of the distribution is relatively short and flat, tall and slender, or somewhere in between.

Figure 2.3 shows the frequency distribution of body temperatures taken over a 24-hour period for a single cat. This distribution is unimodal and symmetric, with the mean, median, and mode all coinciding (at 39.1°C). The range of temperature values is 38.6°C to 39.5°C. Figure 2.4 shows the frequency distribution of heart rate values for the same cat over the same 24-hour period. This distribution is positively skewed, with the mean greater than the median.

2.3.2.3 Factors Influencing the Shape of Frequency Distributions

Actual frequency distributions for many clinical measurements of animal populations change with characteristics such as age, breed, sex, neuter status, plane of nutrition, and, in food-producing animals, stage of production. Consequently, the interpretation of a patient's hematologic or serum biochemical values should take into account these variables.

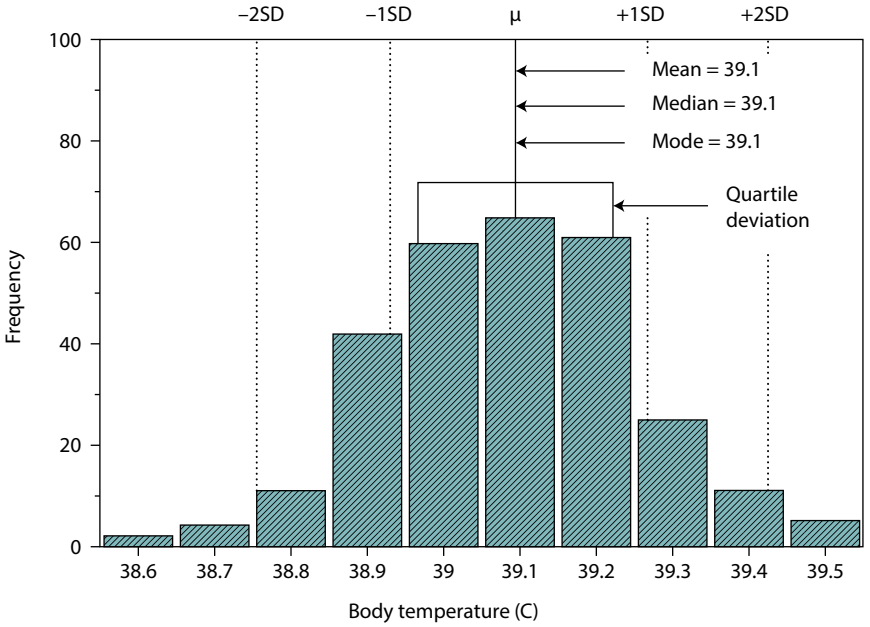

FIGURE 2.3 Frequency distribution of rectal temperature values for a cat over a 24-hour period. (Data courtesy of Dr. R.M. Weigel, College of Veterinary Medicine, University of Illinois. With permission.)

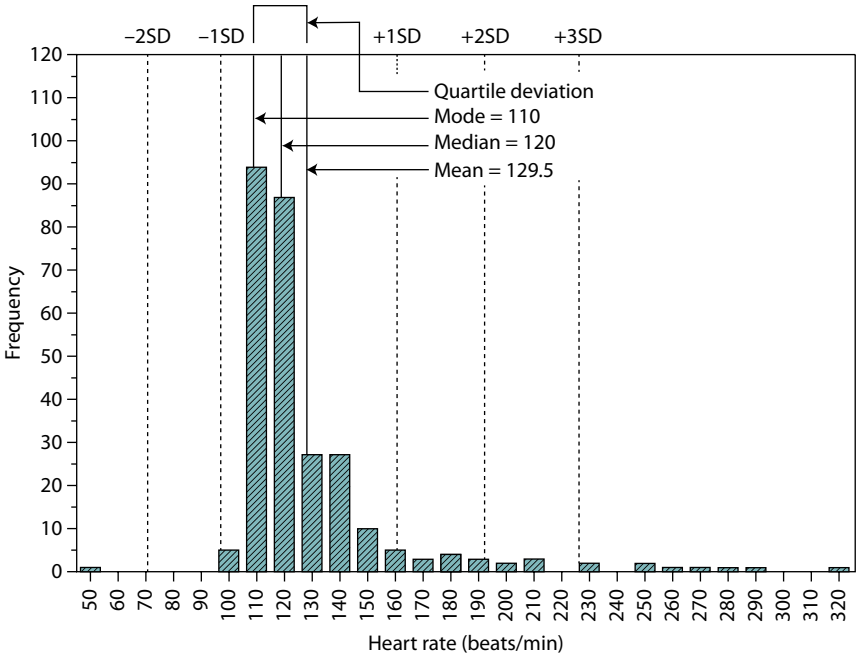

FIGURE 2.4 Frequency distribution of heart rate values for a cat over a 24-hour period. (Data courtesy of Dr. R.M. Weigel, College of Veterinary Medicine, University of Illinois. With permission.)

EXAMPLE 2.6: SHOULD BREED BE TAKEN INTO ACCOUNT WHEN INTERPRETING SERUM BIOCHEMICAL TEST RESULTS IN DOGS?

Background: The serum or plasma biochemical profile is essential in the diagnosis and monitoring of systemic disease in veterinary medicine, but current reference intervals typically take no account of breed-specific differences. Breed-specific hematological phenotypes have been documented in the domestic dog, but little has been published on serum biochemical phenotypes in this species.

Objectives: Chang et al. (2016) conducted a large-scale retrospective analysis of phenotypic diversity of biochemical analytes by age, sex, and breed in the domestic dog, both to assess the need for breed-specific reference intervals and to lay the foundations for the genetic dissection of these traits in the future, ultimately triangulating phenotype, breed, and genetic predisposition.

Study Design: Cross-sectional.

Methods: Serum biochemical profiles of dogs in which all measurements fell within the existing reference intervals were retrieved from a large veterinary database. Serum biochemical profiles from 3045 dogs were retrieved, of which 1495 had an accompanying normal glucose concentration. Sixty pure breeds plus a mixed-breed control group were represented by at least 10 individuals.

Results: All analytes, except for sodium, chloride, and glucose, showed variation with age. Total protein, globulin, potassium, chloride, creatinine, cholesterol, total bilirubin, ALT, CK, amylase, and lipase varied between sexes. Neutering status significantly impacted all analytes except albumin, sodium, calcium, urea, and glucose. Tentative reference intervals were generated for breeds with a distinctive phenotype identified by comparative analysis and

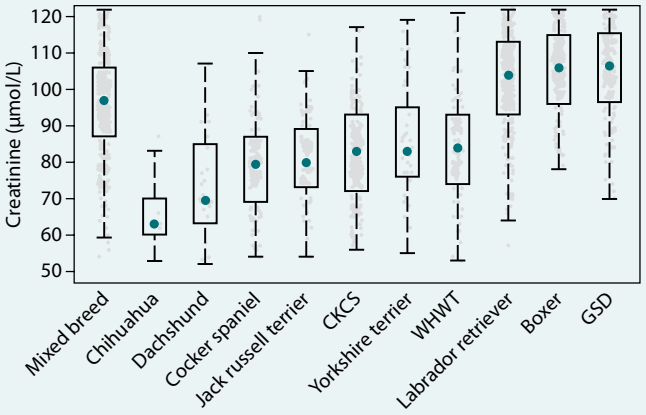

FIGURE 2.5 Combined box-and-whisker/dot plots representing serum creatinine concentrations from breeds showing a significant difference from the mixed breed group. Each small dot represents an individual dog; the boxes show the respective 25th and 75th percentiles, the larger dots median values, and the whiskers the lowest and highest data points still within 1.5 times the interquartile range of the respective lower and upper quartiles. CKCS, Cavalier King Charles Spaniel; WHWT, West Highland White Terrier; GSD, German Shepherd Dog. (From Chang Y-m et al. *PLOS ONE*. 2016;11:117. With permission.)

represented by at least 120 individuals. While many of the absolute differences were small and of questionable clinical significance, some, such as those for serum creatinine concentration (Figure 2.5), used for staging of chronic kidney disease, were notable and would warrant breed-specific reference intervals adjusted for age, sex, and neutering status.

Conclusions and Significance: This is the first large-scale analysis of breed-specific serum biochemical phenotypes in the domestic dog and highlights potential genetic components of biochemical traits in this species. The genetic homogeneity of canine breeds offers a truly unique resource to investigate the genetic basis for homeostatic control of serum biochemical analytes in health, thus ultimately promising to disclose novel therapeutic targets in disease.

FOLLOW-UP QUESTION 2.6

Figure 2.6 depicts the frequency distribution of serum creatinine concentrations in German Shepherd dogs obtained by the authors. How would you describe this data distribution? (*Hint*: See Table 2.3.) See answer 2.6 at the end of this chapter.

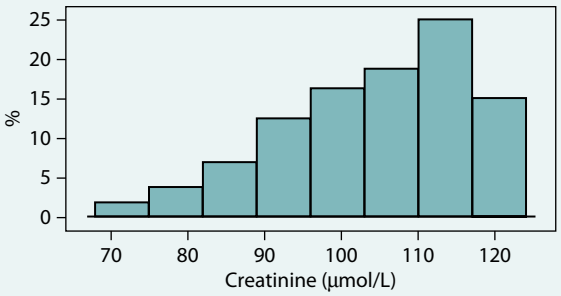

FIGURE 2.6 Frequency distribution of serum creatinine concentration in German Shepherd dogs. (From Chang Y-m et al. *PLOS ONE*. 2016;11:1–17. With permission.)

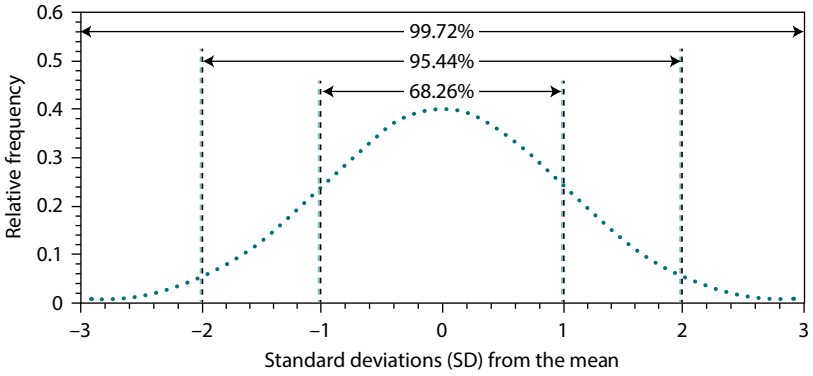

FIGURE 2.7 Percentages under the normal (Gaussian) curve at various standard deviations.

2.3.3 The Normal Distribution

At this point it is important to draw a distinction between the naturally occurring distributions discussed above and the **normal or Gaussian distribution**, the symmetrical bell-shaped curve that is frequently used as the standard that biological data are assumed to fit. The normal distribution (Figure 2.7) is a mathematical or theoretical model that describes the distribution of repeated measurements of the same physical property by the same instrument. The dispersion of these measurements thus represents random variation alone. Because the frequency distribution for many continuous random variables in biology *approximates* a normal distribution, the latter is frequently used as a mathematical or theoretical model for calculating central tendency and dispersion. In clinical epidemiology it is frequently used to calculate the limits of normality.

The mathematical representation of the normal distribution is not discussed here, but some consequences of the mathematical formulation for the shape and other distribution properties of the normal distribution should be mentioned. The normal distribution is unimodal, with the mean equal to the median equal to the mode. It is symmetrical, meaning that within a given number of **standard deviation** (SD) units from the mean, there will be the same proportion of values in the positive direction as in the negative direction. Approximately two-thirds of all values will be within ±1 SD of the mean, approximately 95% of values will be within approximately ±2 SDs from the mean, and approximately 99% of values will be within approximately ±3 SDs from the mean in a normal distribution.

2.4 REFERENCE RANGES AND THE CRITERIA FOR ABNORMALITY

We now come to a crucial point: given the variety of clinical measurements and dispersion inherent in animal data, **how do we determine what is normal and what is abnormal**? The distribution of clinical values among normal and diseased individuals frequently overlaps.

Figure 2.8 depicts the frequency distribution of body temperatures for a group of clinically normal dogs (from Figure 2.2) superimposed on that for dogs exhibiting various signs of respiratory or gastrointestinal disease such as runny eyes and nose, harsh lung sounds, coughing, diarrhea, and lethargy. Not only is the shape of each histogram different, but there is a significant degree of overlapping of normal with abnormal.

> ***When there is no clear division between normal and abnormal, three criteria have proven useful: being unusual, being sick, and being treatable (Fletcher et al., 1996).***

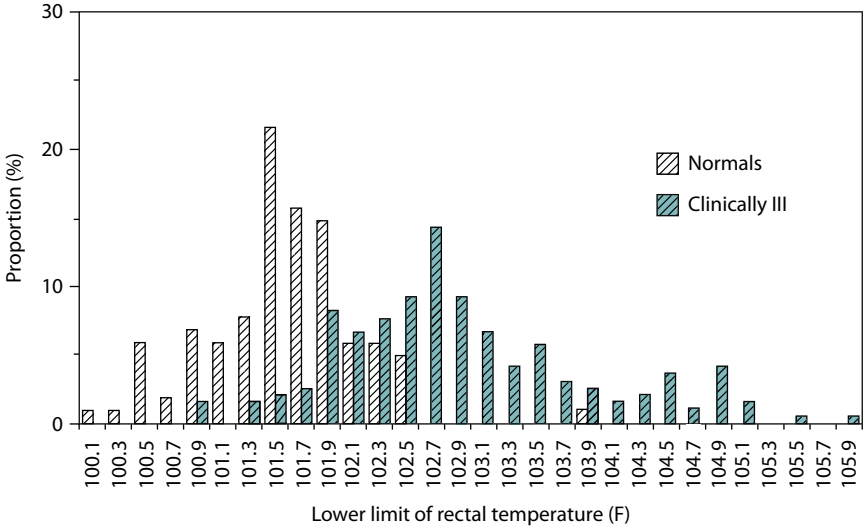

FIGURE 2.8 Frequency distribution of rectal temperatures for clinically normal and clinically ill dogs.

2.4.1 Abnormal as Unusual

The criteria for abnormality may be approached statistically. One approach assumes that normal clinical values exhibit a Gaussian distribution. Thus, if we arbitrarily define the cutoffs (e.g., **critical values**) between normal and abnormal to be the mean ±1.96 SDs, then 95% of the reference values would be within the normal range and 5% outside (2.5% on each end of the distribution). In the example of normal canine body temperatures (Figure 2.2), the mean or average temperature (μ) was 101.6°F with an SD of ±0.6°F. Application of these criteria would yield a maximum normal temperature of 102.8°F.

These criteria are the basis for **two-tailed tests of significance**. This approach is fine if we do not want to specify abnormality as being above or below our normal range, e.g., a *nondirectional hypothesis of normality* (Figure 2.9).

Sometimes a **one-tailed test of significance** is more appropriate, as when we wish to define where fever begins. In this case we are not interested in the bottom of the normal range, but rather the top, e.g., above normal body temperature. The one-tailed approach still defines normal as 95% of reference values, but the 5% abnormals all come from the right-hand side of the bell-shaped

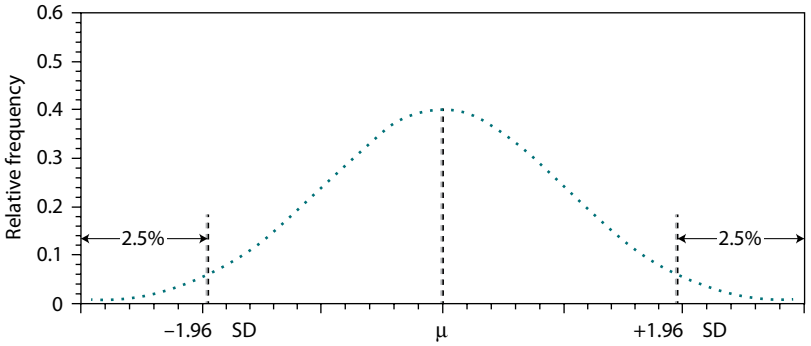

FIGURE 2.9 Mean (μ) and critical values (±1.96 SD) for the 95% confidence interval, under a ±two-tailed test of significance, where abnormality is associated with either high or low values, as blood leukocyte counts.

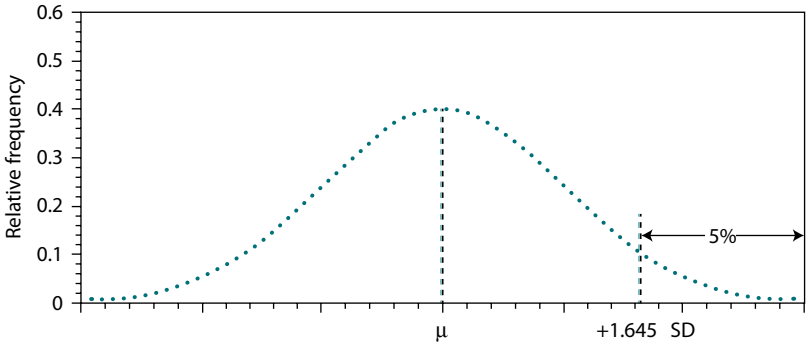

FIGURE 2.10 Mean (μ) and critical values (+1.645 SD) for the 95% confidence interval, under a one-tailed test of significance, where abnormality is associated with high values, as body temperatures.

curve (Figure 2.10). As a result, the normal/abnormal cutoff would be shifted "to the left" (critical value = +1.645 SD), resulting in a more conservative estimate of normal. In this case the one-tailed approach would yield a maximum normal body temperature of approximately 102.6°F.

There are two limitations to the statistical approach to defining normality. First, if we define the normal range as comprising 95% of the reference population, then 5%, or 1 in 20, would fall outside of the normal range. Since the entire reference population was "normal" to begin with, these would be **false positive results** for the condition that we are measuring. The likelihood of false positives increases when multiple test panels are interpreted. As more tests are added to a panel, it becomes more likely that a normal individual will have at least one falsely abnormal result. If we were to extend the normal range to include 99% of the reference population, then the likelihood of classifying truly diseased individuals as normal, e.g., **false negative results,** increases.

The second important limitation to the statistical approach to normality is that mean and SD determinations assume that the data being analyzed follow a Gaussian (i.e., bell-shaped or normal) distribution. The normal distribution represents only random variation, whereas clinical measurements are subject to many other sources of variation. As a result, if distributions of clinical measurements from many individuals resemble normal curves, it is largely by accident. The canine temperature data in Figure 2.2 approximate the normal distribution. Other data, such as serum creatinine concentration in German Shepherd dogs (Figure 2.6), do not. It is often assumed, as a matter of convenience, that clinical measurements are normally distributed.

Before making an assumption of normality, one should determine whether the distribution can in fact be approximated by a normal curve. This may be done simply by constructing a histogram of the data and looking for obvious departures from a normal distribution, such as skewing. A more formal approach would be to perform a **Chi-square goodness-of-fit test**. If the data are not normally distributed, then one could define normal as the 2.5%–97.5% percentile range of the cumulative distribution. This approach is independent of the shape of the distribution curve and provides an attractive alternative for determining critical values. By this method, the cutoff for the upper 2.5% of the normal distribution for canine temperatures would be 102.6°F (see Figure 2.2).

EXAMPLE 2.7: HOW DO THE STATISTICAL APPROACHES TO DEFINING NORMAL REFERENCE RANGES COMPARE?

The canine serum biochemical study described in Example 2.6 and partially depicted (creatinine only) in Figure 2.5 demonstrates the difficulty of defining normal reference ranges in animal populations. Table 2.4 summarizes the results of applying three different statistical methods (normal distribution, percentiles, range) to defining normal/abnormal cutoffs for

TABLE 2.4
Statistical Summary of Central Tendency and Dispersion of Serum Creatinine Values (μmol/L) for Canine Breeds Depicted in Figure 2.5

Breed	N	Mean	SD	95% CI[a]	Median	$X_{0.025}$–$X_{0.975}$[b]	Range[c]
Mixed	242	95.6	14.7	66.9–**124.4**	97	64–120	54–122
Chihuahua	14	66.6	10.5	**46.1–87.2**	63	53–85.7	53–87
Dachshund	34	73.1	14.6	**44.5–**101.7	69.5	52–107	52–107
Cocker Spaniel	98	79.9	13.0	54.4–105.4	79.5	57–107.9	54–120
Jack Russell Terrier	78	81	12.1	57.3–104.7	80	58.7–105	54–115
CKCS[d]	174	82.6	13.7	**55.8–**109.3	83	58.3–108	56–117
Yorkshire	41	85.9	15.4	55.8–116	83	62–116	55–119
WHWT[e]	78	84.4	14.8	55.3–113.5	84	57.7–116	53–121
Labrador Retriever	327	102.4	13.4	76–**128.7**	104	71.6–122	57–122
Boxer	146	105.1	11.0	83.5–**126.6**	106	82.6–122	78–122
German Shepherd	160	104.9	12.3	80.8–**129.1**	107.5	78.9–122	70–122
All Breeds	1392	94	17	60.7–**127.2**	95	60–121	52–122

Source: Chang Y-m et al. *PLOS ONE*. 2016;11:117. With permission.
Note: Values in boldface are outside of actual measured ranges.
[a] 95% confidence intervals were calculated as the mean ± 1.96SD.
[b] 2.5%–97.5% percentile distribution; 95% of measured values fall within these intervals.
[c] Lowest and highest values in the distribution; all measured values fall within these intervals.
[d] Cavalier King Charles Spaniel.
[e] West Highland White Terrier.

serum creatinine concentration in the canine breeds depicted in Figure 2.5. Of particular interest are cases in which limits predicted by the normal distribution (based on mean and standard deviation) fall outside of the actual range of values obtained. Eight of the 12 breed categories exhibit this statistical error, a result of applying **parametric statistics** to a skewed data distribution. Further, the magnitude of the breed effect cannot be overlooked. Dachshunds and German Shepherds exhibited the greatest departure from actual limits.

FOLLOW-UP QUESTION 2.7

Which of the three approaches to defining normal reference ranges summarized in Table 2.4 would you consider most appropriate for use in practice? (*Hint*: See Table 2.3.) See answer 2.7 at the end of this chapter.

The statistical approach to normality is useful in many situations, but in others, different criteria are needed.

2.4.2 Abnormal as Associated with Disease

This approach relies on calling abnormal those findings that are regularly associated with disease, disability, unproductivity, or death. An example might be the different classes of heart murmurs associated with valvular defects, or the "pinging" sound one hears on auscultation of the abdomen

EXAMPLE 2.8: IS THERE A RELATIONSHIP BETWEEN NEWBORN PUPPY IgG CONCENTRATIONS AND THE RISK OF DEATH?

Background: The prevalence of neonatal mortality is high in dogs and not well studied. In most domestic neonates, an appropriate colostrum intake is a key element in the prevention of neonatal mortality.

Objectives: Mila et al. (2014) evaluated the impact of passive immune transfer on puppy mortality, assessed through serum immunoglobulin G (IgG) concentration at 2 days of age. Factors impacting passive immune transfer and the value of an oral immunoglobulin supplementation to prevent it were also analyzed.

Study Design: Randomized controlled clinical trial.

Methods: A total of 149 puppies from 34 litters (12 breeds) within one breeding kennel were included. Blood samples were collected at 2 days of age and colostrum was collected from their dams 1 day after whelping to assay IgG concentration. Puppies were weighed at birth and at 2 days of age for calculation of growth rate. Mortality was recorded until 3 weeks of age. Seventy randomly assigned puppies were orally supplemented with hyper-immunized adult plasma twice within the first 8 hours of life.

Results: Neonatal mortality and weight gain were significantly correlated with puppy IgG concentration at 2 days of age, but not with breed size, sex, supplementation, litter size, or colostrum IgG concentration in a multivariable model with litter as a random effect. The critical threshold between high and low mortality was determined to be 230 mg/dL, calculated as the best result from the equation: test sensitivity + test specificity – 1 (Youden's J index; see Chapter 3 for a discussion of these concepts). Among puppies with IgG concentration ≤230 mg/dL at 2 days of age, 44.4% (12/27) died versus only 4.9% (6/122) of puppies with higher IgG concentrations (Figure 2.11). This relationship is both statistically and clinically significant.

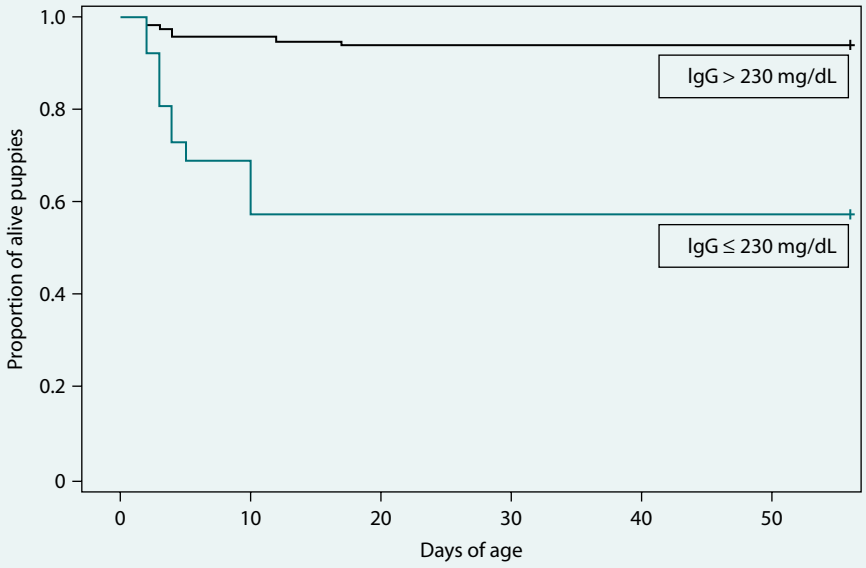

FIGURE 2.11 Kaplan–Meier plot of survival kinetics in puppies above (n = 123) and below (n = 26) the critical threshold of puppy IgG concentration at 2 days of age (230 mg/dL) for higher risk of death ($x^2 = 30.33$; $p < 0.001$). (From Mila H et al. *Prev Vet Med* 2014; 116: 209–213. With permission.)

TABLE 2.5
Risk Factor Assessment for Mortality in Puppies between 2 and 21 Days after Birth Evaluated by a Generalized Linear Mixed Model (n = 149)

Factor	*p* Value	Odds Ratio	95% Confidence Interval	
Breed size	0.386	0.3	0.0	4.8
Sex	0.765	0.8	0.2	3.1
Supplementation	0.851	0.8	0.1	5.0
Litter size	0.382	1.2	0.8	2.0
Colostrum IgG	0.728	1.0	0.9	1.2
Puppy IgG at Day 2	0.018	0.7	0.5	0.9

Source: Mila H et al. *Prev Vet Med* 2014;116:209–213. With permission.

Conclusions and Significance: This study demonstrates that neonatal mortality in puppies is related to the quality of passive immune transfer, and that early colostrum intake should be monitored to reduce neonatal mortality. Oral supplementation with hyper-immunized canine plasma neither decreased risk of mortality nor improved serum IgG concentration at 2 days of age in puppies. No effective alternative source of immunoglobulins is currently available for puppies.

FOLLOW-UP QUESTION 2.8

In addition to IgG concentration, the authors evaluated the relationship between a number of other variables and puppy survival. How many of the other potential risk factors listed in Table 2.5 have a statistically significant association with puppy survival? (*Hint*: See Chapter 9, "Statistical Significance for help.")

of cows suffering from displaced abomasum. This approach is fundamental to the evaluation of diagnostic tests, where the frequency of findings in cases and noncases of a disease is compared. This concept will be discussed more extensively in the next two chapters.

2.4.3 Abnormal as Detectable or Treatable

For some conditions, the level of disease at which intervention is practical may determine whether a particular clinical measurement is considered abnormal. The decision to treat is usually based on evidence from clinical trials. The definition of treatability frequently changes with the accumulation of new knowledge. Consider, for example, parasitism in horses. As the efficacy of anthelmintics for equine strongyles has increased, the egg per gram (EPG) counts tolerated by owners and practitioners have steadily declined. A comparable phenomenon has occurred over the years with drug and chemical residues. As the sensitivity (e.g., absolute sensitivity or detection limits) of assays and instruments has improved, the tolerable level of many substances in animal tissues, fluids, and products has decreased.

In food animal and population medicine, abnormality may be defined as the point at which an intervention (surveillance, vaccination, test and removal, treatment) is economically justified. This point, termed the **economic threshold** or **break-even point**, is dependent on the cost of the intervention and the economic benefit that can be expected, and is usually determined through

a process called **cost-benefit analysis** (discussed in Chapter 14, "The Cost of Disease"). To be effective in these situations, a veterinarian must be knowledgeable in economic analysis as well as in medicine.

REFERENCES

Barber LG and Weishaar KM. Criteria for designation of clinical substage in canine lymphoma: A survey of veterinary oncologists. *Vet Comp Oncol* 2016;14:32–39. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/vco.12086

Campora C, Freeman KP, and Baral Rl. Clinical application of biological variation data to facilitate interpretation of canine and feline laboratory results. *J Small Anim Pract* 2018;59:3–9. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/jsap.12781

Chang Y-m, Hadox E, Szladovits B et al. Serum biochemical phenotypes in the domestic dog. *PLOS ONE* 2016;11:1–17. Full text available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0149650

Falkenö U, Hillström A, von Brömssen C et al. Biological variation of 20 analytes measured in serum from clinically healthy domestic cats. *J Vet Diagn Invest* 2016;28:699–704. Full text available at: https://journals.sagepub.com/doi/full/10.1177/1040638716666602?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed

Fletcher RH, Fletcher SW, and Wagner EH. *Clinical Epidemiology—The Essentials*, 3rd ed. Baltimore: Lippincott Williams & Wilkins; 1996.

Freeman KP, Baral RM, Dhand NK et al. Recommendations for designing and conducting veterinary clinical pathology biologic variation studies. *Vet Clin Pathol* 2017;46:211–220.

Hart BL, Hart LA, Thigen AP et al. The paradox of canine conspecific coprophagy. *Vet Med Sci* 2018;4:106–114. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5980124/

McMillan FD, Serpell JA, Duffy, DL et al. Differences in behavioral characteristics between dogs obtained as puppies from pet stores and those obtained from noncommercial breeders. *J Am Vet Med Assoc* 2013;242:1359–1363.

Mila H, Feugier A, Grellet A et al. Inadequate passive immune transfer in puppies: Definition, risk factors and prevention in a large multi-breed kennel. *Prev Vet Med* 2014;116:209–213.

Mosby's Medical Dictionary, 9th ed. Mosby Elsevier; 2012; St Louis, Missouri.

Owen, L.N., ed., *TNM Classification of Tumours in Domestic Animals*, 1st ed. Geneva, WHO; 1980.

Sebbag L, Kass PH, and Maggs, DJ. Reference values, intertest correlations, and test-retest repeatability of selected tear film tests in healthy cats. *J Am Vet Med Assoc* 2015; 246:426–435.

Seibert P, and Walker Smith N. 2005. Medical Records. *The VSPN Notebook*. Full text available to VIN members at: https://www.vin.com/doc/?id=2990807&pid=49.

Straus JH. Anemia. In, Fenner WR (ed). *Quick Reference to Veterinary Medicine*. Philadelphia: J B Lippincott Co; 1982.

Wells DL. Comparison of two treatments for preventing dogs eating their own faeces. *Vet Rec* 2003;15:51–53.

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 2.1: The C-BARQ questionnaire did not define "breeder," leaving participants to define the term for themselves. However, it's unlikely that the owner of a dog coming from a pet store would select "breeder" as a source, or that the owner of a dog from a breeder would select pet store as a source. The sample of dog owners was self-selected ("convenience sample") and therefore a potential source of bias. Owner-related factors (puppy training, demographic factors, background of owners, reporting bias) might have influenced the result. Other sources of puppies (pound, rescue) that might affect the rankings were not considered.

Answer 2.2: The potential discrepancy between the original (percentage) data and the converted score can be appreciated by considering that owner scores of 26% and 50% (a 24% difference) would both be recorded as 2, whereas an owner assessment of 25% coprophagy (only 1% less than 26%) would be recorded as 1. Application of interval-level (parametric; mean, standard deviation) statistics to the analysis of the converted scores would be misleading. The true magnitude and

statistical significance of the difference between treatments is difficult to assess without comparing the original percentage-based data. See Hart et al. (2018) for a more recent evaluation of risk factors and treatments for canine conspecific coprophagy using non-parametric statistical analysis.

Answer 2.3: Although it is not clear how precisely some of the clinical factors listed in Table 2.2 were measured, we can be reasonably sure that hypercalcemia and fever were determined with instrumentation. The other variables can be considered nominal (categorical) or ordinal. It's important to make this distinction, as tests of statistical significance are designed for particular data types. See Chapter 9, "Statistical Significance for criteria used in the selection of a statistical test."

Answer 2.4: Ideally, tear tests performed 1 week apart on a healthy cat would yield the same result, and the mean of the difference between the two tests would be close to 0 (the horizontal line in Figure 2.1). Meibometry appears to be closest to the mean, although there are several extreme outliers. Poor test-retest repeatability for all assays tested in healthy cats may be attributed to measurement error and suggests that the use of any of these assays for the monitoring of disease progression or efficacy of therapeutic interventions may not be reliable.

Answer 2.5: Population-based reference intervals may be stratified in order to increase their utility, and the effect of stratification is greatest when IND_I is low for the whole group but high for a subgroup. However, cofactors suitable for stratification in cats (such as age, breed, and sex) need to be carefully evaluated before any recommendations can be made.

Answer 2.6: The distribution of creatinine values for German Shepherd dogs depicted in Figure 2.6 can best be described as unimodal, skewed to the left (negatively skewed), with the mean probably to the left of the median. The mode (most frequent value) lies somewhere between 110 and 118 μmol/L, and the range is approximately 68–125 μmol/L. As the preceding values are estimates whose precision is influenced by the width of data intervals on the x-axis used to generate the graph, the actual values may vary.

Answer 2.7: Given the skewed nature and breed dependency of reported creatinine distributions, the best strategy would be to define normal limits based on a 95% percentile distribution for each canine breed (Column b in Table 2.4). Patient values lying outside these limits may be clinically significant and should be monitored for trends. For some breeds, the number of individuals sampled was below the threshold of 120 recommended by the authors. In these cases, normal limits based on "All Breeds" can be used but may result in a greater likelihood of false positive or negative results.

Answer 2.8: The only variable in Table 2.5 with a statistically significant relationship ($p < 0.05$) with puppy mortality is serum IgG concentration at 2 days of age. All of the other variables have a greater than 5% chance of showing an association with puppy mortality by chance alone.

3 Evaluation of Diagnostic Tests

3.1 INTRODUCTION

Diagnostic tests play a major role in medical decision-making. In the clinical setting, the results of a diagnostic test may be used to decide whether to initiate or withhold treatment and, if treatment is chosen, to determine the level of treatment. Diagnostic tests are also applied at the herd level to determine the incidence and/or prevalence of disease within the herd, to identify the cause of a disease process, and sometimes to select those animals that should be culled.

A diagnostic test does not have to be laboratory based, but it should provide information upon which decisions can be made. Test results may be reported using any of the three scales described earlier: nominal, ordinal, or interval. A serologic test, for example, may be interpreted as either positive or negative (nominal), strong or weak positive (ordinal), or reacting up to a given dilution of serum or titer (interval).

A distinction must be made between diagnostic and screening test scenarios. **Diagnostic testing** is used to distinguish animals that have the disease in question from those that have other diseases on the differential list. Diagnostic testing begins with diseased individuals. **Screening** is used for the presumptive identification of unrecognized disease or defect in apparently healthy populations. Screening begins with presumably healthy individuals. The same test, examination, or procedure may be used for either purpose. The distinction is necessary because of the nature of the population used to standardize the test and the effect of disease prevalence on the interpretation of test results.

This chapter discusses how the properties of diagnostic tests are evaluated and expressed. The subsequent chapter presents guidelines, or rules, for their application in medical decision-making. Techniques employed for the evaluation of diagnostic tests are summarized in Table 3.1.

> ***A distinction must be made between diagnostic and screening test scenarios. Diagnostic testing begins with diseased individuals, whereas screening begins with presumably healthy individuals.***

3.2 TEST ACCURACY

Test accuracy is the proportion of all test results, both positive and negative, that are correct. Another term for accuracy is **validity**. Accuracy is often used to express the overall performance of a diagnostic test. Because accuracy answers the question, "What is the likelihood that the test result is correct?" this test property is of great interest in clinical practice.

The accuracy of diagnostic tests falls on a continuum. As a general rule, as tests become more accurate they also become more tedious, invasive, and costly. The choice of simpler tests over more elaborate and accurate diagnostic strategies must be made with the realization that some risk of misclassification exists, which is justified by the feasibility and cost of the simpler tests. The choice of a particular test requires a balance between the risk of making an incorrect diagnosis and the relative cost of false-positive and false-negative results (Theurer et al., 2015). As a result, diagnostic testing is frequently approached in stages, substituting simpler tests for more rigorous ones, at least initially.

TABLE 3.1
Techniques for the Evaluation of Diagnostic Tests

Test Parameter Being Evaluated	How Measured	How Expressed
Validity	2-by-2 table	Sensitivity, specificity, positive and negative predictive values, accuracy
Optimum cutoff	Response-operating characteristic (ROC) curve	Positive/negative cutoff value
Comparison of tests	Fixed cutoff: Pre-test/Post-test curve	Posterior probability ÷ prior probability
	Continuous variable: Response-operating characteristic (ROC) curve	Likelihood ratio at different levels of the test; area under the curve
Clinical utility	True positive rate ÷ false positive rate; false negative rate ÷ true negative rate	Likelihood ratio for a positive or negative test
	Decision analysis[a]	Testing and treatment thresholds

[a] See Chapter 14.

EXAMPLE 3.1: HOW MUCH WEIGHT SHOULD BE PLACED ON CLINICAL SIGNS IN THE DIAGNOSIS OF BOVINE RESPIRATORY DISEASE?

Background: Bovine respiratory disease complex (BRD) is a major cause of illness in dairy calves. The diagnosis of active infection of the lower respiratory tract is challenging on a daily basis in the absence of accurate clinical signs. Clinical scoring systems such as the California scoring system (Aly et al., 2014; Love et al., 2014) are appealing, but were developed without considering the imperfection of reference standard tests used for case definition.

Objectives: Buczinski et al. (2018) used a Bayesian latent class model to update California prediction rules for BRD using thoracic ultrasonography (TUS) as a reference standard. A second objective of the study was to determine the optimal strategy for using this updated test in situations with different clinical settings (likelihood of disease and relative cost of false negative/false positive cases).

Study Design: Cross-sectional (prevalence survey).

Methods: The results of clinical examination and ultrasound findings of 608 pre-weaned dairy calves from 39 herds were used to rank the diagnostic accuracy of clinical findings and recommend how best to incorporate them into a diagnostic protocol for BRD. Clinical signs were assessed by two different individuals in accordance with guidelines for clinical scoring described for the California scoring chart (Aly et al., 2014). A systematic bilateral thoracic ultrasonography was performed on all enrolled calves by one experienced veterinarian and a recently graduated veterinarian who received specific training in TUS.

Results: Two hundred twenty (36.2%) calves showed evidence of lung consolidation ($\geq$1 cm consolidation) on thoracic ultrasonography. Revised weights were assigned to the six clinical signs used in the California scoring system: 20 points for abnormal breathing pattern, 16 for ear drop/head tilt, 16 points for cough, 10 points for the presence of nasal discharge, 7 points for rectal temperature $\geq$39.2°C, and –1 point for the presence of ocular discharge. Using this system, a total of 64 clinical sign permutations are possible, with scores ranging from 0–70. Clinical scores between 9 (83% sensitivity, 69.1% specificity) and 16 (66.9% sensitivity, 82.7% specificity) resulted in the lowest cost of misclassification under low to

average (≤20%) BRD prevalence. The optimal cut-off decreased when expected prevalence of disease and the false negative/false positive ratio increased. A decision tree (see Chapter 14 for an expanded discussion of decision trees) based on the clinical signs described above yielded 64 possible different clinical sign profiles. The decision tree provided an alternative method for predicting the probability (95% credible intervals) of an active infection of the lower respiratory tract.

Conclusions and Significance: Dairy practitioners can use either the updated clinical scoring system, selecting the optimal cut-off based on the clinical context (high vs. low expected prevalence of respiratory problems), or use the results from incorporating the 64 clinical sign profiles into the decision tree. The scoring system could also be used as a screening test before applying other more specific tests such as thoracic auscultation or determination of serum acute phase protein concentrations in calves with abnormal clinical sign profiles. The applications of these clinical signs at the group level would also be helpful in deciding whether a group of calves is at high or low risk of being BRD affected

FOLLOW-UP QUESTION 3.1

What sources of bias might limit the strength (internal validity) and generalizability (external validity) of the authors' conclusions? See Answer 3.1 at the end of this chapter.

3.2.1 The Standard of Validity (Gold Standard)

Ideally, all diagnostic tests should be backed by sound data comparing their accuracy with an appropriate standard. The **gold standard**, sometimes called the "definitive test," refers to the means by which one can determine whether a disease or other condition is truly present. Its function is that of a quality-control device. The gold standard provides the basis for determining the value of diagnostic tests, treatment strategies, and prognoses. In some cases, a simple microbiologic culture or blood smear is sufficient to confirm the presence or absence of disease. In others, more elaborate, risky, and expensive tests must be used, each with its own inherent accuracy.

The gold standard is a quality-control device that provides the basis for determining the value of diagnostic tests, treatment strategies, and prognoses.

Postmortem examination, or necropsy, is often regarded as the ultimate confirmational test. A well-performed necropsy is an instrument of quality control and a supplier of data on disease processes and the accuracy of diagnosis and treatment (Schertenleib et al., 2017; Wäsle et al., 2017). However, some diseases require confirmatory microbiological testing to arrive at a diagnosis (Proaño-Pérez et al. 2011). Further, many disorders cannot be confirmed even at necropsy because they stem from subtle biochemical or neurologic alterations measurable only in the living animal.

3.2.2 Postmortem Examination as a Diagnostic Test

Postmortem examination is used more frequently as a diagnostic tool in veterinary medicine than in human medicine (Dank et al, 2012; Kent et al, 2004). Besides its value as a quality-control device for monitoring the accuracy and interpretation of other diagnostic tests, postmortem examination offers a number of other benefits. When combined with patient history, it can provide information

on the efficacy and toxicity of therapeutic agents, permit the detection of conditions that may have been important but were either clinically inapparent or obscured by the most prominent disease, and help to monitor the influence of environmental factors on physiologic processes. In addition, postmortem examination is a highly effective method for exploring the variable manifestations of animal diseases.

EXAMPLE 3.2: IS ANTE-MORTEM DIAGNOSTIC ACCURACY IN VETERINARY PRACTICE IMPROVING?

Background: Kent et al. (2004) compared necropsy rates and the discrepancy between clinical and pathological diagnoses over a 10-year interval at the University of California, Davis Veterinary Medical Teaching Hospital (VMTH). They reported a decrease in the necropsy rate between 1989 (58.9%) and 1999 (48.3%), while the discrepancy between clinical and pathological diagnoses did not change significantly (39.8% and 37%, respectively).

Objectives: In a follow-up study, Dank et al. (2012) sought to determine the proportion of discrepancies between clinical and pathological diagnoses made during 2009, 1999, and 1989 and whether the proportion of dogs necropsied had changed.

Study Design: Cross-sectional (prevalence survey).

Methods: Medical records of 148 hospitalized dogs that died or were euthanized at the VMTH during 2009 were reviewed. The medical records of the dogs enrolled in this study were reviewed to retrieve the ante-mortem clinical and post-mortem pathologic diagnoses. Clinical and pathological diagnoses were recorded, categorized, and compared to historical controls using a data set of 623 dogs studied in 1989 and 1999.

Results: The proportion of discrepancies was significantly ($p < 0.001$) lower in 2009 (14.9%), compared to both 1999 (37%) and 1989 (39.8%), reflecting improved ante-mortem diagnostic accuracy over the last decade. There was also a significant ($p < 0.001$) decrease in the proportion of necropsies performed during 2009 (21.4%) compared to both 1999 (48.4%) and 1989 (58.9%).

Conclusions and Significance: There was a marked improvement in the ante-mortem diagnosis of patients in 2009 compared with both 1989 and 1999, as evidenced by the decrease in the proportion of discrepancies between the clinical and pathological diagnoses. Despite the decline in the proportion of necropsies performed, the authors consider necropsies a vital tool for teaching, determining the pathological basis of disease, identification of new and emerging diseases, and determining the cause of death in individual animals. The authors caution that their findings may not be generalizable to private companion animal practice because of the nature of cases referred to the VMTH and their clients' support for more thorough diagnostic workups.

FOLLOW-UP QUESTION 3.2

What might account for the improved ante-mortem diagnostic accuracy measured by the investigators over the 1999–2009 time period? See Answer 3.2 at the end of this chapter.

3.3 PROPERTIES OF DIAGNOSTIC TESTS

The performance characteristics of diagnostic tests can be evaluated by using the two-by-two table depicted in Figure 3.1. Data must be obtained for all four cells.

	True Disease Status	
Test Result	**Present**	**Absent**
Positive	(a) True Positive	(b) False Positive
Negative	(c) False Negative	(d) True Negative

- Test sensitivity = a ÷ (a + c)
- Test specificity = d ÷ (b + d)
- Predictive value of a positive test = a ÷ (a + b)
- Predictive value of a negative test = d ÷ (c + d)
- Positive likelihood ratio = sensitivity ÷ (1 − specificity)
- Negative likelihood ratio = (1 − sensitivity) ÷ specificity
- Test accuracy = (a + d) ÷ (a + b + c + d)
- True prevalence = (a + c) ÷ (a + b + c + d)
- Apparent prevalence = (a + b) ÷ (a + b + c + d)

FIGURE 3.1 Diagnostic test outcomes and definitions for a test with binary outcomes (i.e., positive vs. negative results). There are four possible test outcomes: two are correct and two are incorrect. Values for all four outcomes are used to estimate test sensitivity, specificity, predictive values, accuracy, positive and negative likelihood ratios, and true and apparent prevalence.

3.3.1 Sensitivity and Specificity (True Positive and True Negative Rates)

Two special terms are traditionally used to describe the characteristics of a test. **Test sensitivity**, or **true-positive rate**, is defined as the likelihood of a positive test result in patients known to have the disease or condition in question (pT+/D+). Test sensitivity has also been referred to as "**operational sensitivity**" to distinguish it from "**analytic sensitivity**," a term used to express the detection limits of an assay. **Test specificity**, or **true-negative rate**, is the likelihood of a negative test result in patients known to be free of the disease or condition (pT−/D−).

Case series are excellent sources of data on the sensitivity of a particular test or finding.

EXAMPLE 3.3: WHAT ARE THE MOST COMMON FINDINGS ASSOCIATED WITH ELAPID SNAKEBITE ENVENOMATION IN HORSES?

Background: Snakebites can lead to severe clinical illness and death in horses. Many studies of snakebite in the horse originate from North America and describe envenomation by members of the crotalid sub-family (pit vipers, including rattlesnakes). Elapid snakes are the predominant family in Australia, where tiger snakes (*Notechis spp.*), brown snakes (*Pseudonaja spp.*) and black snakes (*Pseudechis spp.*) are most commonly implicated in domestic animal envenomation. Although bites from elapid snakes are often speculated to be a cause of sudden illness and unexpected death in endemic areas, there is a paucity of literature describing elapid snake envenomation in horses, and the evidence for envenomation is sometimes circumstantial.

Objectives: Bamford et al. (2018) sought to describe the presentation, clinical and laboratory findings, treatment, and outcome of horses with a diagnosis of elapid snake envenomation in Australia.

Study Design: Case series.

Methods: The clinical findings and outcome of 52 confirmed cases of elapid snake envenomation in Australia based on medical records (2006–2016) retrieved from several university and private veterinary practices were reviewed. Inclusion criteria comprised one or more of the following: (1) observed snakebite, (2) positive snake venom detection kit (SVDK) result, (3) appropriate clinical response to treatment with antivenom, or (4) supportive post-mortem findings.

Results: Table 3.2 ranks the frequency of reported clinical and pathologic findings. Of 18 urine samples evaluated with a SVDK, only three (17%) were positive. Overall survival was favorable (86%) among 49 horses who received antivenom. Eighteen surviving horses (43%) required more than one vial of antivenom.

TABLE 3.2
Dichotomous Data for Clinical Examination and Clinical Pathology Variables Recorded on Admission in 52 Horses Diagnosed with Elapid Snake Envenomation

	Data Available (n)	Present n (%)
Clinical Examination		
Tachycardia	50	48 (96)
Neuromuscular weakness	52	49 (94)
Altered mentation	52	48 (92)
Muscle fasciculations	52	44 (85)
Recumbency	52	39 (75)
Absent/reduced PLR	40	30 (75)
Mydriasis	47	34 (72)
Tachypnoea	49	33 (67)
Sweating	48	26 (54)
Pyrexia	48	23 (48)
Tongue paresis	30	13 (43)
Dyspnea	50	20 (40)
Dysphagia	15	6 (40)
Pigmenturia	49	15 (31)
Colic signs	52	11 (21)
Hypothermia	48	5 (10)
Clinical Pathology		
Hyperlactatemia	25	20 (80)
Leucocytosis	37	22 (59)
Rhabdomyolysis[a]	40	20 (50)
Hyperfibrinogenemia	25	6 (24)
Azotemia	37	9 (24)
Hemolysis	43	8 (19)

Source: Bamford NJ et al. *Equine Vet J* 2018;50:196–201. With permission.
Abbreviations: PLR, pupillary light response.
[a] Defined as plasma creatine kinase >3000 U/L.

Conclusions and Significance: Test sensitivity data such as these provide useful criteria for ranking diseases on a differential list. For example, tachycardia, neuromuscular weakness and altered mentation (dull, agitated, hyper-responsive) were each present in more than 90% of cases. Thus, if an animal is believed to be suffering from elapid snake bite but does not exhibit one or more of the most frequent signs, it may be wise to consider other diseases on the differential list. Diagnostic suspicion is best confirmed by response to treatment with antivenom.

FOLLOW-UP QUESTION 3.3

Neuromuscular weakness was reported in 94% of cases of equine elapid snake envenomations. If an equine patient presents with neuromuscular weakness, can we be 94% sure that the bite of an elapid snake is responsible? See Answer 3.3 at the end of this chapter.

3.3.2 False Positive and Negative Rates

Two additional rates may be derived from the preceding test characteristics. The **false-positive rate** is the likelihood of a positive test result in patients known to be free of the disease (pT+/D−) and equals (1 – specificity). The **false-negative rate** is the likelihood of a negative result in patients known to have the disease (pT−/D+) and equals (1 – sensitivity).

In summary, sensitivity and the false-negative rate describe how the test performs in patients with a disease, whereas specificity and the false-positive rate describe how the test performs in patients without the disease.

3.3.3 Predictive Values

Although a test's sensitivity and specificity are important properties, clinicians should be more concerned with a test's **predictive value**, i.e., the probability that a test result reflects the true disease status (see Figure 3.1). **Positive predictive value** is the probability of disease in an animal with a positive (abnormal) test result (pD+/T+). **Negative predictive value** is the probability that an animal does not have the disease when the test result is negative (pD−/T−). Whereas sensitivity and specificity can be regarded as absolute properties of a test (with the possible exception described in the example below), predictive values are relative, varying with the likelihood, or **pretest probability**, of disease in the individual being tested. The pretest probability of disease may be based on patient history, the clinician's experience with similar patients, or the prevalence of the disease in the population from which the individual was drawn. For a full discussion of prevalence, see Chapter 5.

Strictly speaking, prevalence of disease cannot influence test sensitivity and specificity in the way that it affects predictive values. However, there are situations in which test sensitivity and specificity may differ between populations of high and low prevalence. For example, the sensitivity of antigen tests for canine heartworm has been shown to increase with increasing worm burdens (Courtney et al., 1988). Courtney and Cornell (1990) have discussed how the distribution of different types and intensities of heartworm infection (patent, immune-mediated occult, unisex occult, immature occult, high and low worm burdens) may differ among canine populations in regions of high and low endemicity or among different classes of dogs, thereby affecting the overall sensitivity of the test. Consequently, test sensitivity based on a study of Florida dogs, where worm burdens are high, may be much higher than one could expect in regions of low endemicity. In a similar fashion, antibody titers to a disease agent may be higher among animals residing in areas with a high prevalence of that disease, thereby affecting the sensitivity and specificity of antibody tests.

3.3.4 The Effect of Prevalence on Predictive Values

Diagnostic tests are used in populations with widely differing disease frequencies. As indicated above, prevalence per se has no effect on test sensitivity or specificity, but predictive values may vary considerably. As the prevalence of infection decreases, the positive predictive value also decreases, but the negative predictive value increases. An easy way to appreciate the effect of prevalence upon predictive values is to consider test results in populations where all or none of the individuals have a disease. In the former case, the predictive value of a positive test would be 100% and that of a negative test would be 0%. In contrast, if no individuals are diseased, then the predictive value of a positive test would be 0% and that of a negative test 100%.

The predictive value of diagnostic results can be improved by selecting more sensitive or specific tests. A more sensitive test improves the negative predictive value of the test (fewer false-negative results). A more specific test improves the positive predictive value (fewer false-positive results). However, because prevalence commonly varies over a wider range than sensitivity or specificity, it is still the major factor in determining predictive value. Improved sensitivity and specificity cannot be expected to result in a dramatic improvement in predictive value.

The decline of the predictive value of a positive test with decreasing prevalence is of special concern in test and removal programs for disease eradication among food-producing animals. Use of a serologic test of low specificity (and therefore low positive predictive value) could result in excessive culling of disease-free individuals from a herd.

The relationship among the above measures of diagnostic test performance can be appreciated from the following report on the development of an enzyme-linked immunosorbent assay (ELISA) for dermatophytosis in cats. It is also an example of the "One Health" concept in veterinary and human medicine.

EXAMPLE 3.4: CAN FELINE RINGWORM BE DIAGNOSED SEROLOGICALLY?

Background: Dermatophytosis (ringworm) is the most common fungal infection in cats worldwide and plays an important role in both animal and human health due to its high zoonotic potential. The most common etiologic agents belong to the genera *Microsporum* and *Trichophyton*. They are highly contagious and readily transmissible to humans, and as many as 50% of people exposed to infected cats develop ringworm lesions. Effective screening is a strong preventive measure. Some breeds of cats serve as asymptomatic carriers and under the right circumstances (crowding, close contact) can serve as a source of outbreaks of dermatophytosis. Fungal culture is the "gold standard" for diagnosis of dermatophytosis. However, fungal culture requires laboratory facilities and results are not readily available. Although a superficial disease, dermatophyte infections induce specific humoral and cellular immune responses.

Objectives: A rapid and accurate screening test for dermatophytosis in cats would contribute to the prevention and control of disease outbreaks. Santana et al. (2018) described the development and diagnostic accuracy of an ELISA for dermatophytosis in cats.

Study Design: Cross-sectional.

Methods: Seventy cats of various ages presented to the veterinary hospital of the University of São Paulo (Brazil) or selected from private shelters of the city of São Paulo were included in the study. Clinical and mycological examinations were performed on each cat. Cats were divided into three groups: S (symptomatic, $n = 20$), AS (asymptomatic, $n = 30$), and N (negative, $n = 20$). Hair samples from all animals were submitted for fungal culture and blood drawn for serological testing. Test antigen was prepared from cultures of six field isolates of

M. canis. The ELISA positive/negative cutoff value was defined as the mean plus two standard deviations above the mean absorbance (optical density; OD) value of the serum samples from culture negative cats (abnormal as unusual; see Chapter 2).

Results: The distribution of ELISA values for culture-positive and negative cats is depicted in Figure 3.2, and test performance at the chosen cutoff summarized in Table 3.3. A significant difference ($p < 0.05$) was found between OD values of *M. canis* positive and negative animals. There was no statistical difference between symptomatic and asymptomatic groups. The ELISA test showed sensitivity of 94% and specificity of 75%. Positive and negative predictive values were 90% and 83%, respectively. Test accuracy, expressed as the area under the receiver operating characteristic curve (ROC; see below), was 92.5%.

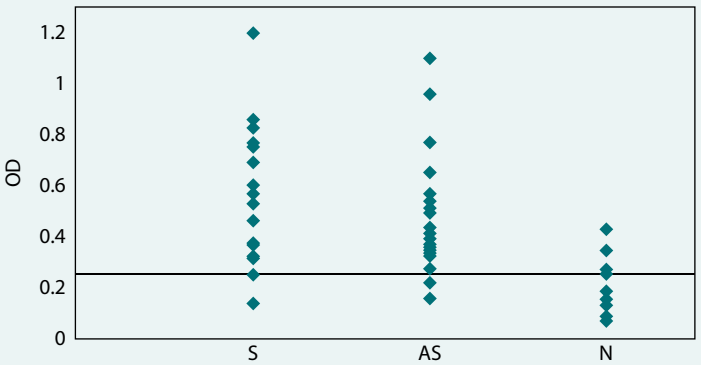

FIGURE 3.2 Reactivity of serum samples from symptomatic (S), asymptomatic (AS), and negative (N) groups in an enzyme-linked immunosorbent assay (ELISA) specific for the *Microsporum canis* antigen. The horizontal line represents the positive/negative cutoff. (From Santana AE et al. *Med Mycol* 2018;56:95–102. With permission.)

TABLE 3.3
Performance Characteristics of an Enzyme-Linked Immunosorbent Assay (ELISA) for the Diagnosis of *Microsporum canis* Infection in Cats

	Reference Standard Test[a]		
Test Under Evaluation	**Positive**	**Negative**	**Total**
Positive	True positive = 47	False positive = 05	52
Negative	False negative = 03	True negative = 15	18
Total	50	20	70

Sensitivity: 94%, Specificity: 75%

Positive predictive value: 90%, Negative predictive value: 83%

Source: Santana AE et al. *Med Mycol 2018*;56:95–102. With permission.

Note: The likelihood ratio for a positive test (LR+) is 3.76 (0.94/0.25).
The likelihood ratio for a negative test (LR–) is 0.08 (0.06/0.75).

[a] Reference standard was based on positive or negative fungal culture.

Conclusions and Significance: The authors concluded that ELISA is a useful tool to reliably distinguish both symptomatic and asymptomatic cats infected with *M. canis* from uninfected animals. The purification of fungal antigens can reduce cross-reactions and increase the test sensitivity and specificity.

FOLLOW-UP QUESTION 3.4

As can be seen in Figure 3.2, some misclassification of culture-positive and negative cats occurred at the chosen ELISA cutoff. Can a cutoff be selected that will not result in any misclassifications? How would changes in the prevalence of infection among sampled cats affect the test parameters summarized in Table 3.3? See Answer 3.4 at the end of this chapter.

Because prevalence commonly varies over a wider range than test sensitivity or specificity, it is still the major factor in determining predictive value. Therefore, improved sensitivity and specificity cannot be expected to result in a dramatic improvement in predictive value.

3.3.5 Likelihood Ratios

The **likelihood ratio** is an index of diagnostic utility that expresses the likelihood that a given finding on the history, physical, or laboratory examination would occur in an animal with, as opposed to an animal without, the condition of interest (Timsit et al., 2018). By "finding," we mean the presence (or absence) of any sign or any of the levels of a laboratory test result, such as an ELISA value. The likelihood ratio is calculated from the same two-by-two table used to calculate other aspects of test performance (Figure 3.1). The **likelihood ratio for a positive test (LR+)** is the ratio of the true-positive rate (pT+/D+) divided by the false-positive rate (pT+/D−), or equivalently, sensitivity/(1 – specificity). The LR+ addresses the question "How much more likely is a diseased animal to have a positive test result, compared with a non-diseased animal?" The **likelihood ratio for a negative test (LR–)** is the ratio of the false-negative rate (pT−/D+) divided by the true-negative rate (pT−/D−), or equivalently, (1 – sensitivity)/specificity. The LR– addresses the question: "How much more likely is a diseased animal to have a negative test result, compared with a healthy animal?" The likelihood ratio does not convey the actual likelihood of disease in an individual, only the likelihood that the test result would occur in an individual with, versus an animal without, the disease.

Interpretation of LRs is intuitive; the larger the LR+, the greater the ability of a diagnostic test to rule in a disease or condition in a patient, whereas the smaller the LR–, the greater the ability of the test to rule out a disease or condition. It is generally accepted that a test with an LR+ > 10 is useful for ruling in a disease, whereas a test with a LR– < 0.1 is useful for ruling out a disease (Timsit et al., 2018). A likelihood ratio of 1 for either a positive or negative test means the test result conveys no useful information.

In the feline ringworm example summarized in Table 3.3, the likelihood ratio for a positive ELISA is 3.76 (0.94/0.25), meaning that a positive test result is 3.76 times more likely to occur in an *M. canis*-infected animal than an uninfected animal. The likelihood ratio for a negative test is 0.08 (0.06/0.75), meaning that the likelihood of a negative ELISA result in an infected versus uninfected animal is 0.08. According to the above criteria, the ELISA described in Example 3.4 above would be most useful for ruling out a diagnosis of *M. canis* infection.

The likelihood ratio offers several advantages over other methods of reporting test performance. Because the likelihood ratio is a ratio of test sensitivity and specificity only, it is unaffected by disease prevalence or probability, making it an especially stable expression of test performance; e.g., it is "portable" from one population or patient to another.

The likelihood ratio is also useful for interpreting test results that fall on a continuum, such as serologic titers or serum biochemical values, where the likelihood of disease increases the more measurements deviate from normal. In this way, test results become more useful for ruling diseases in and out because we are utilizing information that would otherwise be lost if results were simply expressed in terms of a dichotomous positive/negative cutoff. Finally, the likelihood ratio can be used in conjunction with the pretest probability of disease, as suggested by clinical experience or previous test results, to estimate the likelihood of disease given a positive or negative test result. This application of the likelihood ratio will be discussed in the next chapter.

Because the likelihood ratio is a ratio of test sensitivity and specificity only, it is unaffected by disease prevalence or probability, making it an especially stable expression of test performance; e.g., it is "portable" from one population or patient to another.

3.3.6 Accuracy, Reproducibility, and Concordance

Accuracy, reproducibility, and concordance are other terms used to describe diagnostic test performance. As stated above, **accuracy** (or validity) is the proportion of all tests, both positive and negative, that are correct. The numerical limits of test accuracy are its sensitivity and specificity. Accuracy is often used to express the overall performance of a diagnostic test. However, its value is subject to the same constraints as predictive value and is correct only for the population used to standardize the test. As disease prevalence changes, so does the accuracy of the test (except for the special condition where test sensitivity and specificity are equal).

Reproducibility (also known as reliability or precision) refers to the degree to which repeated tests on the same sample(s) give the same result, whereas **concordance** is the proportion of all test results on which two or more different tests agree. An important attribute of test concordance is that as the number of different tests applied to the same sample increases, the likelihood of agreement on all tests decreases.

3.4 INTERPRETATION OF TESTS WHOSE RESULTS FALL ON A CONTINUUM

3.4.1 Trade-Offs between Sensitivity and Specificity

The frequency distribution of test results in normal and diseased animal populations, particularly when measured on an interval scale, forces us to make a trade-off between sensitivity and specificity. Figure 3.3 depicts the distribution of rectal temperatures for the two populations of dogs discussed earlier (see Figure 2.8), with a normal/abnormal (neg/pos) cutoff line superimposed. Because the two distribution curves overlap, moving the cutoff point to the left increases the sensitivity of the test, i.e., the probability of detecting a diseased individual, but decreases the specificity. Moving the cutoff to the right has the opposite effect. There is no way to adjust the cutoff so that sensitivity and specificity are improved at the same time.

The frequency distribution of test results in normal and diseased animal populations, particularly when measured on an interval scale, forces us to make a trade-off between sensitivity and specificity.

3.4.2 Receiver Operating Characteristic Curve

For test results that fall along a continuum, e.g., ELISA OD values (Figure 3.2), test performance can be depicted graphically by plotting a **receiver operating characteristic curve** (also called

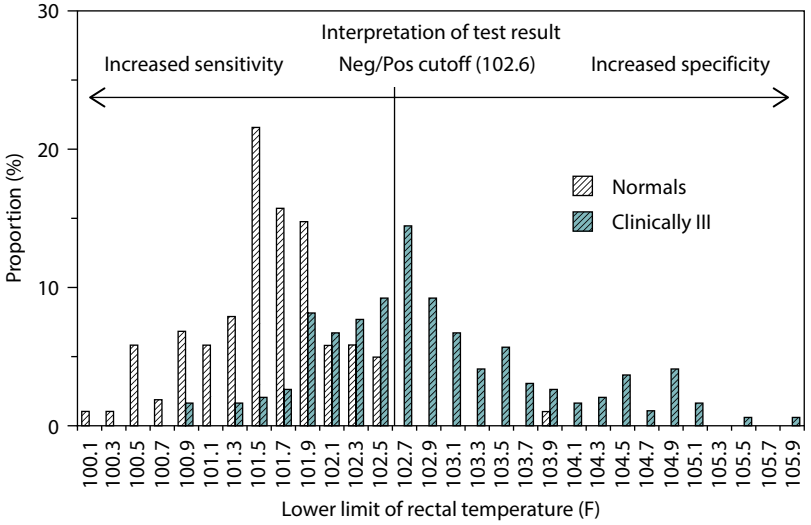

FIGURE 3.3 Frequency distribution of rectal temperatures from normal and clinically ill dogs to demonstrate the effect of moving the negative/positive cutoff on the sensitivity and specificity of the diagnostic test (rectal temperature).

"response operating characteristic curve"), which compares the true-positive rate, or sensitivity, on the vertical axis with the false-positive rate (1 – specificity) on the horizontal axis. ROC analysis is the standard method to demonstrate the co-variation of test sensitivity and specificity and provides a simple method for evaluating a test's ability to discriminate between health and disease over the complete spectrum of operating conditions (cutoffs). ROC analysis can be used to select cutoffs or to compare diagnostic tests. The astute reader will note that the ROC curve is really only a series of likelihood ratios (true positive rate versus false positive rate), using a range of cutoff values as the criteria for test interpretation. Because likelihood ratios are independent of disease prevalence, the ROC curve is a basic tool for the evaluation and application of diagnostic tests (Greiner et al, 2000; Gardner and Greiner, 2006).

An ROC curve for the feline ringworm ELISA data in Figure 3.2 is depicted in Figure 3.4 (Santana et al., 2018). Each point on the ROC curve defines a set of operating characteristics for the test based on sensitivity and specificity. Tests that discriminate well approach the upper-left corner of the ROC curve. The ROC curve for less discriminatory tests fall closer to the diagonal (solid line) running from lower left to upper right. The diagonal line reflects test values that are uninformative, e.g., where the true positive rate equals the false positive rate, e.g., reliance on test results is no better than tossing a coin. The area under the ROC curve (AUC) provides a measure of overall test accuracy and represents the ability of a discriminator (in this case a positive test result) to distinguish an affected from a non-affected individual. According to guidelines proposed by Swets (1988), the AUC can be used to distinguish among non-informative (AUC $= 0.50$), less accurate ($0.50 < \text{AUC} \leq 0.70$), moderately accurate ($0.70 < \text{AUC} \leq 0.90$), highly accurate ($0.90 < \text{AUC} < 1.00$), and perfect tests (AUC $= 1.00$). By these criteria, the feline ringworm ELISA, with and AUC of 0.925, would be considered a "highly accurate" test.

3.4.3 Two-Graph Receiver Operating Characteristic Analysis

A disadvantage of ROC analysis is that it is not possible to read the cutoff value for a selected combination of sensitivity and specificity directly from the ROC plot. **Two-graph receiver operating**

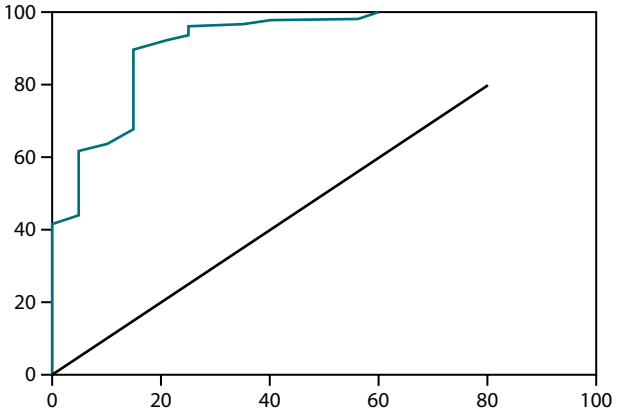

FIGURE 3.4 Receiver operating characteristic (ROC) curve obtained in standardizing the serological enzyme-linked immunosorbent assay (ELISA) for feline ringworm with 70 serum samples (dermatophytosis, other mycosis, other non-fungal diseases, and healthy individuals) using the *Microsporum canis* crude antigen. True positive rate (test sensitivity) is plotted on the vertical axis and false positive rate (1-test specificity) on the horizontal axis. The diagonal line reflects test values that are uninformative, e.g., where the true positive rate equals the false positive rate, e.g., reliance on test results is no better than tossing a coin. The area under the ROC curve (AUC) represents the accuracy of the ELISA; the value is 0.925. (From Santana AE et al. *Med Mycol* 2018;56:95–102. With permission.)

characteristic (TG-ROC) analysis solves this problem. In TG-ROC, test sensitivity and specificity are plotted as separate dependent variables against cutoff values as the independent variable (Greiner et al., 1995). See Example 3.5 (below) for an example of two-graph ROC evaluation of diagnostic tests.

3.4.4 Selecting a Cutoff

Positive/negative cutoffs are used to simplify the diagnostic process by defining the level of a test result that is required to establish or reject a diagnosis. In defining the optimal cutoff, one strives to reduce the consequence of false-negative and/or false-positive test results. Ideally, the choice of a positivity criterion can be likened to a "business decision" that includes a consideration of (1) the distribution of results in two different populations—normal patients and patients with disease, (2) the prevalence or likelihood of disease in the population from which individuals being tested come, and (3) the cost of false-positive and false-negative test results.

> ***Tests of low sensitivity increase the likelihood of false-negative results, whereas tests of low specificity increase the likelihood of false-positive test results.***

The most direct approach is to select the cutoff resulting in the lowest total number of diagnostic errors (false-positive diagnoses plus false-negative diagnoses). The actual prevalence of disease must be known or estimated. At a disease prevalence of about 50%, the optimum cutoff is the point on an ROC curve closest to the upper left-hand corner, where test sensitivity and specificity are maximized, e.g., (sensitivity + specificity – 1) attains its highest value. This point is referred to as **Youden's index** (Greiner et al, 2000; Gardner and Greiner, 2006; Rossi et al, 2018). In TG-ROC analysis, this corresponds to the point where the sensitivity/specificity lines intersect.

EXAMPLE 3.5: HOW ACCURATE IS RESPIRATORY ENDOSCOPY FOR THE DIAGNOSIS OF EQUINE RESPIRATORY DISEASE?

Background: The diagnosis of respiratory disease in horses can be challenging. Diagnosis is usually based on a combination of history (epistaxis related to exercise, poor performance), clinical signs (cough, nasal discharge, fever) and the results of respiratory endoscopy. Visible mucus in the trachea and increased neutrophilia in tracheal wash (TW) and bronchoalveolar lavage fluid (BALF) are common findings in equine lower respiratory tract disease.

Objectives: Rossi et al. (2018) estimated the sensitivity (Se) and specificity (Sp) of TW and bronchoalveolar lavage (BAL) neutrophilia in detecting respiratory disease in a population of client-owned horses with and without signs of respiratory disease.

Study Design: Cross-sectional.

Methods: Retrospective data from 154 horses of various breeds that had been subject to TW and BAL sampling at rest during 2009–2015 were used. The horses were divided into three groups based on the clinical signs and endoscopy findings. Group 1 ($n = 33$) included clinically healthy control horses without signs potentially related to respiratory disease that had been sampled by TW and BAL methods for research purposes with the owners' informed written consent. Group 2 ($n = 79$) included horses that had at least one of the following signs: cough, nasal discharge, epistaxis, fever, or poor performance, and did not belong to either Group 1 or Group 3. Group 3 ($n = 42$) included horses that had both cough and nasal discharge at the time of the examination and/or the highest mucus score during endoscopy that confirmed respiratory disease. Currently accepted cutoff values for neutrophil percentage of >20% in TW and >5% in BALF were considered abnormal. Group 1 and 3 data were compared using two-graph receiver operating characteristic curves to evaluate diagnostic performance of TW and BALF and select the best neutrophil percentage cutoff.

Results: The superior ability of TW to differentiate diseased from healthy horses is apparent from histograms of neutrophil percentage distributions (Figure 3.5a,b) and two-graph ROC analysis (Figure 3.6a,b). It is apparent from a comparison of the distribution of neutrophil percentages for diseased versus control horses (Figure 3.5a,b) that the risk of misclassification is much less for TW versus BALF. The area under the curve for the ROC for TW was 0.884,

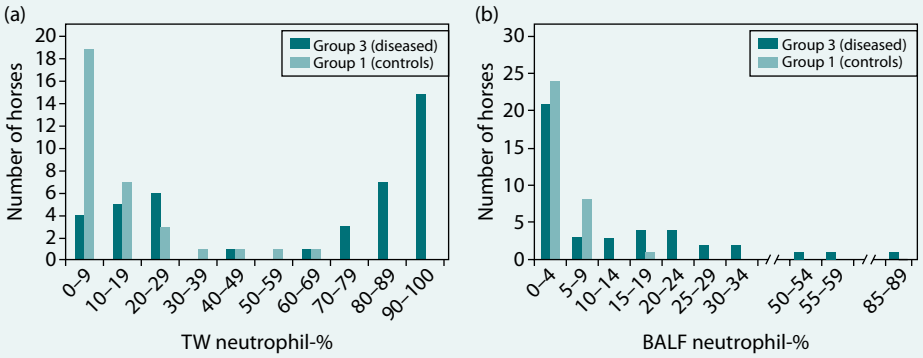

FIGURE 3.5 The tracheal wash [TW; (a)] and bronchoalveolar lavage fluid [BALF; (b)] neutrophil percentage histograms among Group 1 and Group 3 horses. Group 1 included clinically healthy control horses without signs potentially related to respiratory disease (n = 33), and Group 3 included horses that had both cough and nasal discharge and/or highest mucus scores confirming respiratory disease (n = 42). (From Rossi H et al. *Front Vet Sci* 2018;5:61. With permission.)

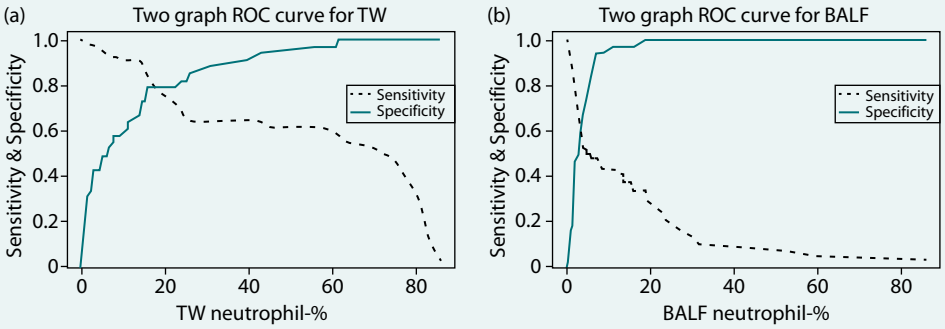

FIGURE 3.6 Two-graph receiver operating characteristic (ROC) curves with sensitivity and specificity plotted against the neutrophil percentage of tracheal wash [TW; (a)] and bronchoalveolar lavage fluid [BALF; (b)] for horses in Group 1 and Group 3. See Figure 3.5 for group definitions. (From Rossi H et al. *Front Vet Sci* 2018;5:61. With permission.)

indicating a moderately accurate test. In comparison, the AUC for the ROC of BALF was 0.685, indicating a less accurate test. Both Se and Sp values were optimized (Youden's *J* value; test sensitivity + test specificity – 1) for TW at a cutoff of 17.7%, yielding approximately 80% Se and Sp in the two-graph ROC curve for TW neutrophil percentage (Figure 3.6a). The optimal cutoff for BALF based on ROC analysis was 7%, yielding approximately 55% Se and Sp (Figure 3.6b).

Conclusions and Significance: The authors concluded that TW was a more sensitive and specific method for diagnosing equine airway inflammation in their patient population, but that further studies with other cell types and in other populations are warranted to determine the best sampling method for individual horses. The choice of an optimal cutoff may also be influenced by the number and relative cost of diagnostic errors.

FOLLOW-UP QUESTION 3.5

What are the possible sources of bias in this study? See Answer 3.5 at the end of this chapter.

3.5 COMPARISON OF DIAGNOSTIC TESTS

3.5.1 Tests with Fixed Cutoffs

The accuracy of a test with a fixed positive/negative cutoff is dependent upon the test's sensitivity, specificity, and the likelihood (pretest probability) of disease in the patient or population. Although test accuracy varies with test sensitivity and specificity, the range of variability is much less than that of a test's positive or negative predictive values. Whereas a test's accuracy cannot be less than its sensitivity or specificity, predictive values can range from 0% to 100%. Despite its variability, accuracy is a convenient way to compare the overall performance of diagnostic tests, particularly if a common value for pretest probability is used in the comparison.

3.5.2 For Test Results That Fall on a Continuum

The evaluation of tests whose results fall on a continuum differs from that for tests with a fixed cutoff because there is no predetermined normal versus abnormal cutoff for the test result.

Using **ROC analysis** (Greiner et al, 2000; Gardner and Greiner, 2006), in which AUCs are calculated over the range of each test's operating conditions, the magnitude and statistical significance of differences among tests can be compared and superior tests identified. Another advantage of ROC analysis is that it can be used for diagnostic tests with outcomes measured on ordinal, interval, or ratio scales.

EXAMPLE 3.6: WHAT IS THE BEST METHOD FOR DETERMINING FAILURE OF PASSIVE TRANSFER OF IMMUNITY (FPTI) IN CALVES?

Background: Calves are born with negligible plasma concentrations of immunoglobulins and must ingest an adequate volume of good-quality colostrum during the first 24 hours after birth to acquire passive immunity via the active uptake of maternal immunoglobulin G (IgG) across small intestinal epithelial cells. Inadequate absorption of colostral IgG_1 is termed failure of passive transfer of immunity and is associated with increased mortality and morbidity in calves' first 6 months of life.

Objectives: Zakian et al. (2018) compared the diagnostic performance of five indirect methods for diagnosing FPTI in Holstein calves at 48 hours of age.

Study Design: Cross-sectional.

Methods: An observational study was performed using a convenience sample of 160 Holstein heifer calves from a herd of 9000 lactating dairy cows in Iran sampled during July of 2015. Serum was harvested at 48 hours of age, and FPTI was assessed using a digital Brix refractometer for total solids measurements, and digital refractometry and the biuret method to measure serum total protein (STP) concentrations. Serum gamma glutamyl transferase activity was measured with an automated analyzer, and serum IgG was measured with the zinc sulfate turbidity test and an enzyme-linked immunosorbent assay. Diagnostic test performance was compared with that of the reference method (FPTI defined as a serum total IgG concentration <10 g/L). Test performance was evaluated using the area under the receiver operating characteristic curve and test sensitivity, specificity, and positive likelihood ratio at the optimal test cut point, and by calculating the kappa coefficient (a measure of the level of agreement between tests).

Results: A serum digital Brix percentage of <7.8% and an STP concentration of <52 g/L measured using digital refractometry were the best methods to identify calves with FPTI (Figure 3.7). Curves for both measures were either very close to or superimposed on the upper-left corner axes of the ROC, indicating extremely high test performance. The STP concentration measured with digital refractometry was 0.1 g/L lower than that measured with the biuret method.

Conclusions and Significance: The digital Brix refractometer and the digital refractometer provide accurate and clinically useful methods for identifying dairy calves with FPTI. In this study, the excellent performance of the Brix refractometer was likely due to the use of a fixed sample volume (200 μL) and a uniform sample temperature at the time of measurement.

FOLLOW-UP QUESTION 3.6

This study was conducted with a convenience sample (animals selected as they became available) of 160 Holstein heifers from one dairy. How might this affect the generalizability (external validity) of the authors' conclusions?

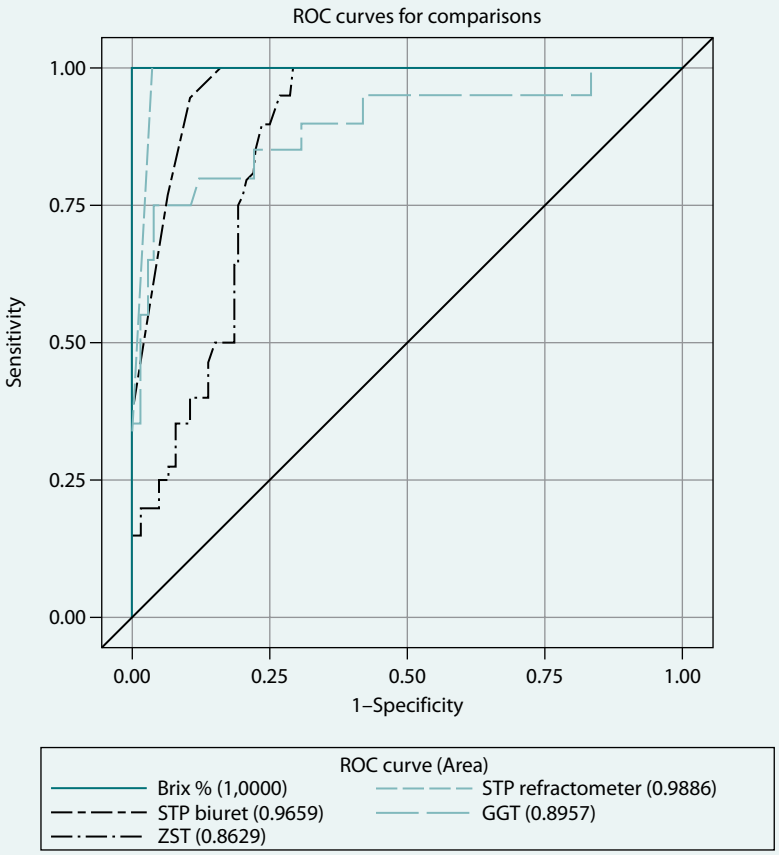

FIGURE 3.7 Receiver operating characteristic (ROC) curves of the five indirect methods for diagnosing failure of transfer of passive immunity (FPTI) in 160 Holstein heifer calves. The area under the ROC curve (AUC) is presented for each method. The optimal cut point for diagnosing FPTI for each method was: digital Brix refractometer at <7.8%; serum total protein concentration (STP) measured using a digital refractometer at <52 g/L; STP concentration measured using the biuret method at <53 g/L; serum gamma-glutamyltransferase (GGT) activity at <815 U/L; and serum zinc sulfate turbidity (ZST) at <14.6 g/L. (From Zakian A et al. *Vet Clin Pathol* 2018;47:275–283. With permission.)

3.6 SOURCES OF BIAS IN THE EVALUATION OF DIAGNOSTIC TESTS

3.6.1 Relative versus True Sensitivity and Specificity

Many times it is not possible to determine the true disease status of animals used for test standardization. However, the "relative" sensitivity and specificity of a diagnostic test can be estimated by comparing test results with those obtained using an accepted standard test that has been in use for many years. This approach might be used by a private practitioner to compare a heartworm serodiagnostic test with the traditional modified Knott's test in client-owned dogs. When there is no gold standard, the comparison of overall performance of one test relative to another is a measure of **concordance** rather than accuracy. Comparisons of the relative accuracy of one test over another are valid only when the true health status of test animals can be determined.

3.6.2 The Spectrum of Patients

Test sensitivity and specificity must be determined in the appropriate population. To establish a test's efficacy for ruling out a diagnosis, sensitivity should be examined in a broad range of patients with the disease. Similarly, to rule in a disease, a test's specificity should be established in a broad range of patients without the disease.

The challenge in the diseased group is to discover whether (and when) the test yields false-negative results. The diseased group should include individuals covering the spectrum of clinical and pathologic findings and those with complications that might yield false-negative results. The distribution of infection stages in the population used to evaluate the test can also affect measurements of test performance.

The challenge in the comparison group is to determine whether (and under what circumstances) the test yields false-positive results. When evaluating a screening test, apparently healthy animals should be used as the nondiseased group. If the same test were to be evaluated in a diagnostic testing scenario, the nondiseased group should consist of animals that do not have the disease for which the test is being evaluated but have other diseases that compete with the disease of interest in the differential diagnosis.

EXAMPLE 3.7: HOW ACCURATE ARE POINT-OF-CARE TESTS FOR DIFFERENTIATING ACUTE PANCREATITIS FROM OTHER CAUSES OF ACUTE ABDOMEN IN THE DOG?

Background: Acute pancreatitis (AP) is an important disease of dogs, with variable and nonspecific clinical signs such as abdominal pain, vomiting, and diarrhea. These clinical signs are also present in conditions such as septic peritonitis or intestinal obstruction, which require specific and timely intervention. Traditional diagnostic methodologies, such as total serum lipase and amylase, have poor sensitivities and specificities for the diagnosis of AP in dogs.

Objectives: Haworth et al. (2014) evaluated the diagnostic performance of two point-of-care serological tests for pancreatic lipase: (1) the SNAP canine pancreatic lipase (cPL) test and (2) the specific canine pancreatic lipase (Spec cPL) test in dogs presenting with acute abdominal disease.

Study Design: Cross-sectional.

Methods: Thirty-eight client-owned dogs that presented to a university teaching hospital emergency center between March 2009 and April 2010 with acute abdominal disease (abdominal pain, vomiting, abdominal distension, or diarrhea), and with a known final diagnosis, were included in the study. Dogs were retrospectively assigned to two groups: dogs with AP (Group 1, $n = 11$) as their primary disease, and dogs with confirmed disease other than AP (Group 2, $n = 27$) based on a complete diagnostic workup that included abdominal ultrasound. Final diagnoses for Group 2 dogs included 12 distinct disease syndromes that should be on the differential list with acute pancreatitis. Estimates of test sensitivity and specificity were based on the ability of the tests to correctly identify dogs in Groups 1 or 2, respectively. Test accuracy was based on the ability of each test to correctly assign dogs to either Group 1 or 2. Test agreement was estimated using McNemar's test and expressed as the kappa (κ) coefficient.

Results: The sensitivity and specificity for SNAP cPL were 82% (9/11 dogs of Group 1) and 59% (16/27 dogs of Group 2), respectively. The sensitivity and specificity for Spec cPL were 70% (7/10 dogs of Group 1) and 77% (20/26 dogs of Group 2), respectively. Accuracy of the SNAP and Spec cPL in arriving at a correct diagnosis was 66% (25/38) and 75% (27/36), respectively. The agreement between SNAP and Spec cPL (cPL $\geq$ 200 μg/L) for the entire study population was $\kappa = 0.78$.

Conclusions and Significance: The authors conclude that SNAP cPL and Spec cPL results may yield a "false positive" diagnosis of pancreatitis in up to 40% of dogs presenting with acute abdominal disease. There was good overall agreement between SNAP cPL and Spec cPL.

FOLLOW-UP QUESTION 3.7

This study was designed to evaluate the ability of two point-of-care diagnostic tests to discriminate between acute pancreatitis and other conditions that might cause acute abdominal disease in dogs.

a. How do you think the tests would have performed if clinically normal dogs had been used as controls?
b. Given the uncertainties of these tests, what can be done to improve their use in the diagnosis of acute pancreatitis?

See Answers 3.7a and b at the end of this chapter.

3.6.3 Bias in Associating Test Results with Disease

Several forms of bias may occur when the status of a test as positive or negative, and the status of disease as present or absent, are not made independently (Ransohoff and Feinstein, 1978, O'Connor and Evans, 2007).

Work-up bias occurs when the results of a test affect the subsequent clinical workup needed to establish the diagnosis of a disease. If a diagnostic test yields a positive result, we are more likely to pursue the diagnosis, increasing the probability of detecting the disease if it is really present. On the other hand, a negative test result may cause us to limit follow-up testing, increasing the probability of missing the disease, if present.

Review bias occurs when the results of a test affect the subjective review of the data that establish the diagnosis. For example, a positive serologic test result may affect the subjective interpretation of thoracic radiographs used to support a diagnosis of occult heartworm disease.

Incorporation bias occurs when the diagnostic test being evaluated, or a related test, is also used to support the diagnosis of the disease. As a result, the case for their use in ruling out disease would be weakened.

3.7 STATISTICAL SIGNIFICANCE

Journal articles may report that a diagnostic test was able to detect a "statistically significant difference" between diseased and comparison groups. However, the magnitude of this difference may not be great enough to be clinically relevant at the level of the individual patient. In some cases, statistical significance is achieved only by using relatively large numbers of animals. If smaller numbers are used, a statistically significant difference may not occur.

REFERENCES

Aly SS, Love, WJ, Williams, DR et al. Agreement between bovine respiratory disease scoring systems for pre-weaned dairy calves. *Anim Health Res Rev* 2014;15:148–150. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4301228/

Bamford NJ, Sprinkle SB, Cudmore LA et al. Elapid snake envenomation in horses: 52 cases (2006–2016). *Equine Vet J* 2018;50:196–201.

Buczinski S, Fecteau G, Dubuc J et al. Validation of a clinical scoring system for bovine respiratory disease complex diagnosis in preweaned dairy calves using a Bayesian framework. *Prev Vet Med* 2018;156:102–112.

Courtney CH and Cornell JA. Evaluation of heartworm immunodiagnostic tests. *J Am Vet Med Assoc* 1990;197:724–729.

Courtney CH, Zeng QY, and Bean ES. Sensitivity and specificity of the Dirochek heartworm antigen test for immunodiagnosis of canine dirofilariasis and a comparison with other immunodiagnostic tests. *J Am Anim Hosp Assoc* 1988;24:27–32.

Dank G, Segev G, Moshe D et al. Follow-up study comparing necropsy rates and discrepancies between clinical and pathologic diagnoses at a veterinary teaching hospital: 2009 versus 1989 and 1999. *J Small Anim Pract* 2012;53:679–83.

Gardner IA and Greiner, M. Receiver-operating characteristic curves and likelihood ratios: Improvements over traditional methods for the evaluation and application of veterinary clinical pathology tests. *Vet Clin Pathol* 2006;35:8–17.

Greiner M, Pfeiffer D, and Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000;45:23–41.

Greiner M, Sohr D, and Gobel P. A modified ROC analysis for the selection of cut-off values and the definition of intermediate results of serodiagnostic tests. *J Immunol Methods* 1995;185:123–32.

Haworth MD, Hosgood G, Swindells KL et al. Diagnostic accuracy of the SNAP and Spec canine pancreatic lipase tests for pancreatitis in dogs presenting with clinical signs of acute abdominal disease. *J Vet Emerg Crit Care* 2014;24:135–143. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/vec.12158

Kent MS, Lucroy MD, Dank G et al. Concurrence between clinical and pathologic diagnoses in a veterinary medical teaching hospital: 623 cases (1989 and 1999). *J Am Vet Med Assoc* 2004;224:403–406.

Love WJ, Lehenbauer TW, Kass PH et al. Development of a novel clinical scoring system for on-farm diagnosis of bovine respiratory disease in pre-weaned dairy calves. *PeerJ* 2014;2:1–25. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898311/

O'Connor A and Evans RB. Critically appraising studies reporting assessing diagnostic tests. *Vet Clin North Am Small Anim Pract* 2007;37:487–497.

Proaño-Pérez F, Benitez-Ortiz W, Desmecht D et al. Post-mortem examination and laboratory-based analysis for the diagnosis of bovine tuberculosis among dairy cattle in Ecuador. *Prev Vet Med* 2011;101:65–72.

Ransohoff DF and Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926–930.

Rossi H, Virtala A-M, Raekallio M et al. Comparison of tracheal wash and bronchoalveolar lavage cytology in 154 horses with and without respiratory signs in a referral hospital over 2009–2015. *Front Vet Sci* 2018;5:61. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5879091/

Santana AE, Taborda, CP, Severo, JS et al. Development of enzyme immunoassays (ELISA and Western blot) for the serological diagnosis of dermatophytosis in symptomatic and asymptomatic cats. *Med Mycol* 2018;56:95–102. Full text available at: https://academic.oup.com/mmy/article/56/1/95/3066110

Schertenleib TI, Pospischil A, Hässig M et al. Comparison of clinical and pathological diagnoses in cats and dogs. *J Comp Pathol* 2017;156:217–234.

Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–1293.

Theurer ME, White BJ, and Renter DG. Optimizing feedlot diagnostic testing strategies using test characteristics, disease prevalence, and relative costs of misdiagnosis. *Vet Clin North Am Food Anim Pract* 2015;31:483–493.

Timsit E, Léguillette R, White BJ et al. Likelihood ratios: An intuitive tool for incorporating diagnostic test results into decision-making. *J Am Vet Med Assoc* 2018;252:1362–1366.

Wäsle K, Pospischil A, Hässig M et al. The post-mortem examination in ruminants and its possible benefit to ruminant clinical medicine. *J Comp Pathol* 2017;156:202–216.

Zakian A, Nouri M, Rasooli A et al. Evaluation of 5 methods for diagnosing failure of passive transfer in 160 Holstein calves. *Vet Clin Pathol* 2018;47:275–283.

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 3.1: *Internal validity:* In lieu of a "gold standard" (necropsy), the authors used thoracic ultrasonography, described by the authors as an "imperfect test." Ultrasonography is not 100%

accurate in the diagnosis of BRD, and misclassification may adversely affect estimates of the validity of clinical signs.

External validity: Only female dairy calves were studied. Thus, the generalizability of the study for other groups of cattle such as feedlot calves remains to be determined. The criteria used for assignment of clinical signs may differ from criteria used by producers and dairy practitioners in the field (inter-observer variability), and thus may influence the validity of clinical scoring.

Answer 3.2 The authors hypothesize that the observed improvement in ante-mortem diagnostic accuracy may be due to advances in veterinary medicine, such as increased specialization, changes in the clinical and pathological faculty over the 20-year period, advances in diagnostic imaging, improved endoscopic and laparoscopic skills, and advancements in laboratory services (e.g., improved laboratory methods, novel molecular tests, and immunostaining of both cytologic and biopsy samples).

Answer 3.3 This is a question of test specificity rather than test sensitivity. Although 94% of horses suffering from elapid snakebite exhibited neuromuscular weakness (high test sensitivity), according to the CONSULTANT Online Veterinary Diagnostic Support System (http://consultant.vet.cornell.edu), neuromuscular weakness can also be a manifestation of at least 231 other equine diseases (e.g., low test specificity). Confirmation of snakebite requires a combination of history and other findings consistent with snakebite.

Answer 3.4 Because the distribution of OD values for culture-positive and culture-negative cats overlap, it is impossible to select a cutoff that does not result in some degree of misclassification. If the prevalence of *M. canis* infection in the population being tested were to change, there would be no change in test sensitivity and specificity, but predictive values would change. The positive predictive value would increase with increasing prevalence of infection and negative predictive value would decrease. The opposite would occur should prevalence decrease.

Answer 3.5 The authors suggest a number of factors that could bias the results. First and foremost, there is no "gold standard," e.g., the diagnostic tests that were being evaluated were also used to classify subjects as diseased or normal. Additionally, as these were client-owned animals, varying stable conditions may have contributed to differences in disease manifestation. Finally, gender and breed differences, and lung and microscopy sampling bias might have influenced the results. Regardless of these caveats, the study represents a more thorough examination of the performance of TW and BALF cytology for the diagnosis of equine respiratory disease.

Answer 3.6 The authors conclude that their study had external validity because the herd genetics and milk production of the dairy are typical for dairy herds in Iran, and the study design was focused on diagnostic test comparison and methods comparison. The study also had adequate statistical power (see Chapter 9) to answer the research questions of interest.

Answer 3.7a In this study the control (non-AP) group consisted of dogs suffering from other causes of acute abdominal disease. Some of these conditions may have indirectly caused a rise in serum pancreatic lipase. Had clinically normal dogs been used as controls, the specificity of both tests would most likely have risen but would have given a false impression of the clinical utility of the tests.

Answer 3.7b A multiple test strategy may significantly improve the utility of these tests in differentiating acute pancreatitis from other causes of acute abdomen in the dog. See the next chapter for a more thorough discussion of multiple testing.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

4 Use of Diagnostic Tests

4.1 INTRODUCTION

The utility of diagnostic tests depends in part on the way in which they are used. Probably the worst approach to medical diagnostics is to perform every conceivable test on a patient in the hope that something will "show up." This would be a waste of hospital and patient resources and would needlessly expand rather than reduce the differential list. Indiscriminate testing at the herd level tends to reduce the predictive value of tests and can lead to unnecessary culling in disease eradication programs. Chapter 3 dealt with the "nuts and bolts" of diagnostic tests. This chapter focuses on strategies that can be used to increase the efficiency of the diagnostic process.

> ***Probably the worst approach to medical diagnostics is to perform every conceivable test on a patient, in the hope that something will "show up."***

4.2 CALCULATION OF THE PROBABILITY OF DISEASE

In the previous chapter, a number of indices of test performance were discussed. These include test sensitivity, specificity, predictive values, likelihood ratios, etc. Although these indices are useful for expressing how well a test performs individually and in comparison with other tests, they do not directly answer the most fundamental question arising from their use: *"What is the likelihood of disease **in this individual** given a positive or negative test result?"* The following section describes a number of ways for answering this question.

4.2.1 From a Two-by-Two Table

The likelihood of disease for a given test result (known as the **post-test probability** of disease) can be estimated directly from the two-by-two table used to evaluate the test's performance. The **likelihood of disease given a positive test result** is, by definition, the test's positive predictive value. The **likelihood of disease given a negative test result** is equal to 100 – negative predictive value (when expressed as a percentage). There is an important caveat to using a two-by-two table in this way. As predictive values are very sensitive to the prevalence (or pretest probability; see below) of disease, it is important that the true prevalence of disease embodied in the two-by-two table (see Figure 3.1) be representative of the population from which the individual being tested was drawn.

4.2.2 Use of Bayes' Theorem

The mathematical relationship among pretest and post-test probabilities and test results was described more than 200 years ago in **Bayes' theorem**. From uncertain beginnings, Bayes' Theorem has become an integral part of decision making affecting all realms of daily life. Readers uncomfortable with statistics and probability theory may find *Bayes' Theorem Examples: A Visual Introduction for Beginners* by Dan Morris helpful.

For epidemiologists, Bayes' theorem provides a theoretical framework for the calculation of post-test probabilities from information that we already know (a priori) about the implications

of a diagnostic test. Using Bayesian analysis, the post-test probability of disease given a *positive* test equals

$$\frac{\text{True-positives}}{\text{All positives}} = \frac{\text{pD} \times \text{Sensitivity}}{\text{pD} \times \text{Sensitivity} + [(1 - \text{pD}) \times (1 - \text{Specificity})]}$$

and the post-test probability of disease given a negative test equals

$$\frac{\text{False-negatives}}{\text{All negatives}} = \frac{\text{pD} \times (1 - \text{Sensitivity})}{\text{pD} \times (1 - \text{Sensitivity}) + [(1 - \text{pD}) \times (\text{Specificity})]}$$

In these equations, the pretest probability of disease (pD), test sensitivity and test specificity must be expressed as a proportion (rather than a percentage).

4.2.3 Use of the Likelihood Ratio to Calculate Post-Test Probabilities

4.2.3.1 Conversion between the Probability of Disease and the Odds of Disease

Regardless of the scale used to report test results (positive/negative or level of a test result), the way in which the likelihood ratio is used to estimate the likelihood of disease is the same. The basic mathematical relationship is represented by the equation:

$$\text{Pretest odds} \times \text{Likelihood ratio (LR)} = \text{Post-test odds}$$

Because this equation is based on the odds of disease, we need to convert disease probability to odds and back again. The conversion between probability of disease and odds of disease is basically a question of converting a rate (probability) to a ratio (odds), and vice versa. In a rate, the numerator is also included in the denominator. Thus, if the prevalence (or probability) of canine heartworm infection is 20%, then 1 *in* 5 dogs (or 0.20 in 1) is infected. In a ratio, the numerator is not included in the denominator. In the above example, the ratio of infected to uninfected dogs would be 1 *to* 4 (or 0.25 to 1). The relationship between probability and odds of disease is expressed mathematically as:

$$\text{Odds of disease} = \frac{\text{Probability of disease present}}{1 - \text{Probability of disease present}}$$

$$\text{Probability of disease} = \frac{\text{Odds of disease}}{\text{Odds of disease} + 1}$$

Thus, if the probability of heartworm infection is 20%, then the odds of heartworm infection would be $0.2 \div 0.8 = 0.25$ (to 1). If the odds of heartworm infection are 0.25 (to 1), then the probability of heartworm infection is $0.25 \div 1.25 = 0.20$.

4.2.3.2 Calculation of the Post-Test Probability of Disease

To illustrate how the likelihood ratio can be used to estimate post-test probabilities, let us return to the use of an ELISA test for diagnosis of *Microsporum canis* infection in cats (Chapter 3; Table 3.3). At the optimal cutoff (an absorbance or optical density value of 0.25), the likelihood ratio for a positive test was 3.76, and for a negative test, it was 0.08. Table 4.1 depicts the results for a positive and negative test result, assuming a pretest probability of infection of 71% (prevalence of culture-positive cats in Table 3.3).

A positive test result would increase the likelihood of *M. canis* infection from 71% to 90%. A negative test result would decrease the likelihood of infection from 71% to 17%. Whenever possible, it is best to express the likelihood ratio for each level of a test result, rather than above or below an

TABLE 4.1

Use of the Likelihood Ratio to Estimate the Post-Test Probability of *Microsporum canis* Infection in Cats with a Positive or Negative ELISA Result

Test Result	Pretest Prob. of Disease	Pretest Odds of Disease	Likelihood Ratio	Post-Test Odds of Disease	Post-Test Prob. of Disease
Positive test result (ELISA $\geq$ 0.35)	0.71	2.5 $\times$	3.76 =	9.40 =	0.90
Negative test result (ELISA $<$ 0.35)	0.71	2.5 $\times$	0.08 =	0.20 =	0.17

Source: Adapted from data in Table 3.3. The pretest probability of infection = 71%, the prevalence of culture-positive cats in the test population.

arbitrary cutoff. For example, in the *M. canis* study, the actual ELISA OD values of all 70 cats ranged from 0.07 to 1.20 (Chapter 3; Figure 3.2). If likelihood ratios were available for each of these values, then the probability of disease could be calculated for any OD value. This information is lost if test results are simply reported as positive or negative based on an arbitrary cutoff.

4.2.3.3 A Nomogram for Applying Likelihood Ratios and Bayes' Theorem

Fagan (1975) offered a solution to Bayes' theorem in the form of a nomogram, a variation of which is depicted in Figure 4.1. The nomogram effectively depicts the relationship among the pretest and post-test probabilities of disease and the likelihood ratio. The pretest and post-test odds have also been

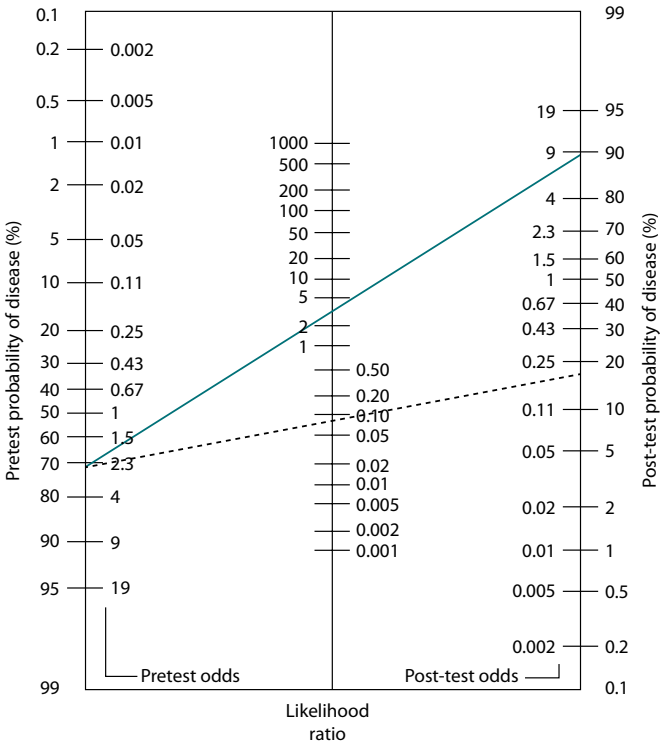

FIGURE 4.1 A nomogram for applying likelihood ratios and Bayes' theorem to the estimation of the post-test probability of disease. Solid and dashed lines represented positive and negative test results (respectively) for an ELISA test for *Microsporum canis* infection in cats, summarized in Table 4.1. (Adapted from Fagan, TJ, *N. Engl. J. Med. (letter)*, 293, 257, 1975. With permission.)

included in the nomogram to help clarify the relationship between probability and odds of disease. Although not as precise as the formulas discussed earlier, the nomogram provides a simple method for visualizing the post-test probability of disease from the pretest probability for any level of a test result.

The utility of the nomogram for estimating post-test probabilities of disease can be appreciated through an analysis of the *M. canis* test data from Table 4.1. Straight lines, representing reported test performance, are superimposed on the nomogram. The solid line corresponds to a positive *M. canis* test result, while the dashed line corresponds to a negative test result. Both lines originate from the vertical axis at the left side of Figure 4.1 at a pretest probability of 71%. They then transect the scale on the center line corresponding to likelihood ratios for positive and negative test results. The post-test probabilities of disease for each test result are the points where the lines intersect the vertical axis on the right side of the graph, 90% for a positive test and 17% for a negative test.

4.2.3.4 Estimating Post-Test Probability of Disease from the Magnitude of a Test Result

For test results that fall on a continuum, there is a correlation between the magnitude of the test result and the likelihood (probability) of disease. This provides the opportunity to express the post-test probability of disease from test results more precisely than is possible through the use of an arbitrary cutoff.

EXAMPLE 4.1: HOW CAN THE DIAGNOSTIC UTILITY OF α_1-ACID GLYCOPROTEIN (AGP) FOR FELINE INFECTIOUS PERITONITIS (FIP) BE IMPROVED?

Background: Feline infectious peritonitis (FIP) is a potentially fatal disease and a common cause of death in young cats. It is caused by feline coronavirus (FCoV). FCoV infection is very common in cats, usually causing only mild gastrointestinal clinical signs such as inappetence, vomiting, and diarrhea. However, up to 10% of FCoV infections result in the fatal form of the infection, FIP (Tasker, 2018). The diagnosis of FIP is often challenging and no curative treatments are currently available. The FCoV antibody titer in an individual animal is of limited use in distinguishing cats with FIP (Tasker, 2018). Serum α_1-acid glycoprotein is now widely used in diagnostic profiles for FIP, but serum AGP levels may be elevated in inflammatory disorders other than FIP. This lack of specificity has limited the clinical utility of serum AGP as a diagnostic test for FIP.

Objectives: Saverio et al. (2007) evaluated the diagnostic potential of AGP for FIP when serum concentrations are expressed on a continuous numerical scale and interpreted using the likelihood ratio (LR) approach and Bayes' theorem.

Study Design: Cross-sectional.

Methods: Serum AGP concentrations from 162 cats of known FIP status (58 cases and 104 non-cases) were used to estimate continuous LRs, which in turn were used to show how Bayes' theorem can be used to estimate the post-test probability of FIP from any serum concentration of AGP.

Results: A receiver-operating characteristic curve was generated to confirm the discriminating power of serum AGP levels for FIP (Figure 4.2a). The discriminating power of AGP is further illustrated in Figure 4.2b, which plots the LR response to increasing concentrations of AGP, which can be used to estimate the likelihood ratio directly from any measured serum AGP concentration, and thereby directly estimate the probability of FIP. Further, the relationship between serum AGP concentrations and LRs can be used in conjunction with Bayes' formula (or the nomogram in Figure 4.1) to calculate the post-test probability of FIP for any value of AGP. The pretest probability (or likelihood) of FIP can be clinically staged based on epidemiological and clinicopathological data (very low, low, medium, high) and used to estimate the post-test probability of disease for any AGP concentration (Table 4.2).

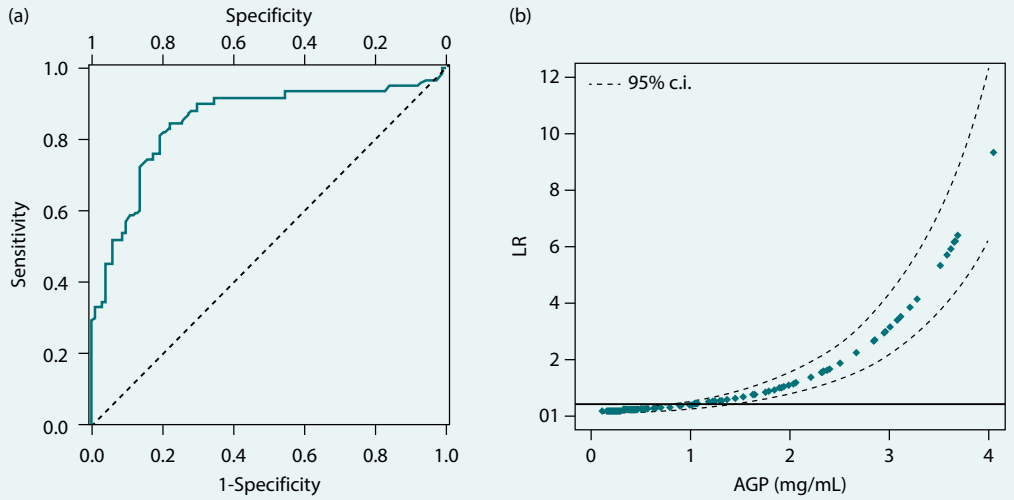

FIGURE 4.2 (a) Receiver-operating characteristic curve of α_1-acid glycoprotein (AGP) for the diagnosis of feline infectious peritonitis (FIP): the area under the curve is 0.85 (SE 5 0.034, 95% confidence interval 5 0.78–0.92), thus confirming that AGP increases have a good discriminating power for FIP. (b) Continuous likelihood ratios and their 95% confidence intervals for any given value of α_1-acid glycoprotein. (From Saverio P et al. *J Vet Diagn Invest* 2007;19:266–272. With permission.)

TABLE 4.2

Post-Test Probability of FIP at different AGP Levels in Four Hypothetical Scenarios Characterized by Different Pretest Probabilities of FIP

Epidemiologic Factors	Clinical Information	FCoV Serology	Pretest Probability	Post-Test Probability (%) for AGP Values (mg/mL) of					
				0.5	1	1.5	2	2.5	>3
Adult pet cat, living alone	Clinically healthy	Negative	1% (very low)	0.23	0.39	0.65	1.09	1.81	>3.02
Adult cat living in a FCoV-endemic cattery	Jaundice, no other clinical or laboratory findings consistent with FIP	Positive	20% (low)	5.38	8.74	13.89	21.36	31.39	>43.52
Young cat, singly housed	Fever, neurological signs, and lymphopenia	Negative	50% (medium)	18.53	27.69	39.21	52.07	64.66	>75
Young cat, living in a FCoV-endemic cattery	Fever, jaundice, hyperproteinemia, decreased A:G ratio, and ascites due to protein-rich, yellowish abdominal effusions	Negative	80% (high)	47.63	60.5	72.07	81.29	87.98	>92.5

Source: Saverio P et al. *J Vet Diagn Invest* 2007;19:266–272. With permission.
Abbreviations: FIP, feline infectious peritonitis; AGP, α_1-acid glycoprotein; FCoV, feline coronavirus.

Conclusions and Significance: The comparison of serum AGP levels in the different groups and the analysis of the ROC curve confirmed that serum AGP is a powerful discriminating marker for FIP. The Bayesian approach demonstrated that when the pretest probability of FIP is high, based on history and clinical signs, moderate serum AGP levels (1.5–2 mg/mL) can discriminate cats with FIP from others, while only high serum AGP levels (>3 mg/mL) can support a diagnosis of FIP in cats with a low pretest probability of disease. The diagnostic utility of AGP might otherwise have been reduced if an arbitrary positive/negative cutoff had been used.

FOLLOW-UP QUESTION 4.1

What is missing from Table 4.2 that was used to calculate the post-test probabilities of FIP from serum AGP concentrations? See Answer 4.1 at the end of this chapter.

4.2.4 Use of Post-Test Probabilities in Medical Decision-Making

Besides its inherent value as an expression of the likelihood of disease, the post-test probability can be used to rank the likelihood of diagnoses on a differential list or to reconcile a series of test results, where the post-test probability after one test becomes the new pretest probability for the next test. This sequential approach works as long as certain conditions are met. Most importantly, the test(s) must either be conditionally independent (i.e., the sensitivity and specificity of the second test must not depend on the results of the first) or all conditional dependencies must be explicitly described (i.e., the probability of the second test being positive, given both disease *and* a positive result for the first test).

4.3 MULTIPLE TESTS

Diagnoses are seldom made on the basis of a single test. **Multiple testing**, e.g., using tests in combination, is common in the veterinary hospital and in the field. The interpretation of multiple test results depends on the sequence in which they are conducted and the way in which their results are integrated. This section discusses the principles by which multiple tests are interpreted. Table 4.3 summarizes the factors to be considered in ordering and interpreting multiple tests.

4.3.1 Parallel Testing

In **parallel testing**, two or more different tests are run on a patient or herd, usually at the same time. A positive diagnosis requires that only one of the test results be positive. A common example of parallel testing is the initial screening of outpatients during vaccination clinics. Typically, a careful physical examination is conducted and the temperature, pulse, and respiratory rate are recorded. The degree of overlap in the distribution of these parameters among normal and sick animals is considerable.

Diagnostic tests are usually done in parallel when rapid assessment of the patient's condition is necessary, as in emergency or hospitalized patients, or emergency-care patients where the health status of the patient will determine whether a subsequent procedure can be performed. *The net effect of parallel testing is to ask the patient to prove that it is healthy.*

Parallel testing is particularly useful when the clinician is faced with the need for a very sensitive test but has available only two or more relatively insensitive ones. By using the tests in parallel, the net effect is a more sensitive diagnostic strategy with a higher negative predictive value (see Example 4.2 below). On the other hand, specificity and positive predictive value are lowered. Only

TABLE 4.3
Aspects of Multiple Test Strategies

	Multiple Test Strategy		
Considerations	**Parallel Testing**	**Serial Testing**	**Herd Retest**
Interpretation of test strategy	Positive diagnosis requires only one positive test result	Positive diagnosis requires that all test results be positive	Positive diagnosis requires only one positive test result
Effect of test strategy	Increase sensitivity	Increase specificity	Increase sensitivity at the herd level
Greatest predictive value	Negative test sequence	Positive test sequence	Negative test sequence
Application	Rule out a disease	Rule in a disease	Rule out a disease
Purpose; clinical setting	Rapid assessment of individual patients; vaccination clinics, emergencies	Time not crucial; avoid excessive testing of groups of animals; test and removal programs	Time not crucial; test and removal programs
Comments	Useful when there is an important penalty for missing a disease, i.e., false-negative results	Useful when there is an important penalty for false-positive results	Useful when there is an important penalty for missing a disease, i.e., false-negative results

animals that have negative results on all tests are considered truly free of disease. The price is evaluation or treatment of some patients without disease.

A disadvantage of parallel testing is that as the number of tests included in the testing strategy increases, the risk of false positive diagnoses also increases. If the clinician orders enough tests, a new abnormality will be discovered in virtually all healthy patients. The reason is obvious if we recall that the "normal range" of values is usually defined to include 95% of the normal population. Thus, as the number of unrelated tests performed in parallel increases, the chance that a patient will be normal on all tests decreases. On the other hand, normal results on parallel tests increase the likelihood that the patient is truly normal. Parallel testing is usually used on an individual-patient basis rather than on groups of animals, such as litters, kennels, or herds.

4.3.2 Serial Testing

In **serial testing**, two or more different tests are run on a patient or herd, but all test results must be positive for a positive diagnosis to be made. *The net effect is to ask the patient to prove that it is truly affected by the condition being sought*. Serial testing maximizes specificity and positive predictive value, but lowers sensitivity and negative predictive value. We can be more confident in positive test results but run an increased risk that disease will be missed.

Serial testing may be used during the course of a diagnostic workup, where rapid assessment of patients is not required, or when some of the tests are expensive or risky (these tests being employed only after simpler and safer tests suggest the presence of disease). CONSULTANT (http://consultant.vet.cornell.edu), an online veterinary diagnostic support system, employs a serial test strategy to generate a differential list from findings entered by the user. As the number of findings increases, the number of diseases that share all of the findings decreases, resulting in a progressively shorter differential list.

Serial testing is also an integral part of disease control programs. Typically, screening tests are followed by confirmatory tests of positive herds or animals to reduce the likelihood that healthy animals are needlessly culled from the herd (see Example 4.2 below).

Serial testing should not be confused with **paired sampling**, in which the same test is performed on sequentially collected samples to detect changes in test results over time. Either

decreases or increases in measured results may be sought. For example, paired sampling may be used to reduce the likelihood of false-positive results due to time-related phenomena peculiar to particular patients. Examples are colostral antibody or vaccination titers, which can be distinguished from immune responses to actual infection by their tendency to drop off over time. During an outbreak investigation, paired sampling may be performed to identify a potential causative agent. In this case, an increased antibody titer to a specific pathogen in paired sera collected during and following a disease outbreak provides presumptive evidence for its involvement in the outbreak.

The net effect of parallel testing is to ask the patient to prove that it is healthy, whereas the net effect of serial testing is to ask the patient to prove that it is truly affected by the condition being sought.

EXAMPLE 4.2: HOW DO COMBINATION TEST STRATEGIES FOR CANINE LEISHMANIASIS AFFECT TEST ACCURACY?

Background: Canine visceral leishmaniasis (CVL) is a potentially fatal zoonotic disease caused by the protozoan parasite *Leishmania infantum*, which is transmitted by several species of sandflies. In Brazil, human and canine visceral leishmaniasis are endemic and have undergone a process of urbanization over the past few decades. Domestic dogs have been suggested as the main reservoir of *L. infantum* in urban areas, and euthanasia of serologically positive dogs has been proposed as a means of reducing human infections. Inappropriate culling of dogs due to test inaccuracies has been the subject of much controversy.

Objectives: In an effort to improve test accuracy Arruda et al. (2016) compared the performance of an enzyme-linked immunosorbent assay and an indirect immunofluorescent antibody (IFA) test for *L. infantum* when performed individually and in combination.

Study Design: Cross-sectional.

Methods: Samples from 98 cases and 1327 noncases were included, yielding a prevalence of infection in the study population of 6.88%. The reference (gold) standard was based on the direct visualization of parasites in histological sections, immunohistochemistry, or isolation of the parasite in culture.

Results: The results of interpreting the tests individually and in combination are summarized in Table 4.4. Individually, both tests presented sensitivity of 91.8% and 90.8%, and specificity of 83.4 and 53.4%, for the ELISA and IFA, respectively. When tests were used in parallel combination (a positive result on either test interpreted as a positive test overall), sensitivity was 99.2%, while specificity dropped to 44.8%. When used in serial combination (ELISA as a screening test followed by IFA as the confirmatory test), sensitivity decreased to 83.3% and specificity increased to 92.5%. The serial testing approach thus improved specificity with only a moderate loss in sensitivity in the study population. A sensitivity analysis was conducted over the range of prevalences (1%–67%) likely to be encountered in endemic areas of Brazil (Figure 4.3). At low to moderate prevalence levels ($\leq$40%), optimal results for positive and negative predictive values were achieved when the tests were interpreted in sequence (serially), and reduced the likelihood of euthanatizing uninfected dogs. At high prevalence levels (>40%), where emphasis is placed on identifying and removing potential reservoirs of infection, a parallel testing strategy might be preferable.

TABLE 4.4

Accuracy of Enzyme-Linked Immunosorbent Assay (ELISA) and Indirect Immunofluorescence Assay (IFA), Individually and Combined in Sequence or Parallel, in Serum Samples of Dogs from Areas Endemic for Visceral Leishmaniasis in Brazil

	ELISA[a] %	IFA %	Serial Testing[b] %	Parallel Testing %
Accuracy	**(95% CI)**	**(95% CI)**	**(95% CI)**	**(95% CI)**
Sensitivity	91.8 (86.3–97.3)	90.8 (84.2–97.5)	83.3 (75.6–91.0)	99.2 (75.6–91.0)
Specificity	83.8 (81.8–85.7)	53.4 (51.6–55.1)	92.5 (87.1–97.9)	44.8 (87.1–97.9)
PPV	29.6 (24.5–34.7)	12.6 (7.1–18.1)	44.9 (34.4–55.1)	11.7 (5.0–18.1)
NPV	99.3 (98.8–99.8)	98.7 (98.2–99.3)	98.7 (96.4–100)	99.8 (98.9–100)

Source: Arruda MM et al. *Mem Inst Oswaldo Cruz* 2016;111:168–173. With permission.

Abbreviations: CI, confidence interval; NPV, negative predictive value; PPV, positive predictive value.

[a] Data in this column were reported previously in de Arruda et al. (2013).

[b] ELISA followed by IFA.

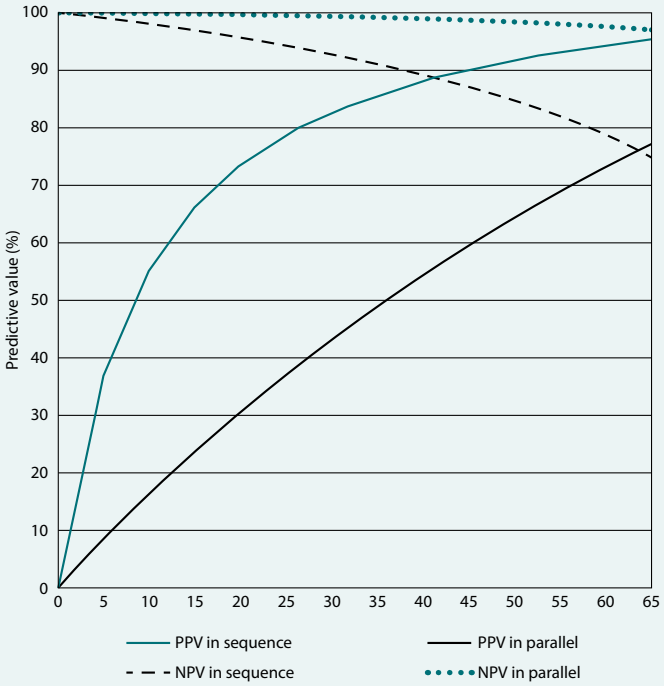

FIGURE 4.3 Sensitivity analysis of positive (PPV) and negative (NPV) predictive values of test combinations for visceral leishmaniasis diagnosis for enzyme-linked immunosorbent assay (ELISA) and indirect immunofluorescence assay (IFA) (Bio-Manguinhos) in serial (ELISA followed by IFA) or parallel combinations, according to prevalence levels of canine visceral leishmaniasis. (From Arruda MM et al. *Mem Inst Oswaldo Cruz* 2016;111:168–173. With permission.)

Conclusions and Significance: These results demonstrate that it is possible to improve the diagnostic accuracy of testing for CVL through the use of a combined serological testing approach and that the accuracy of the two tests combined in sequence reached reasonable sensitivity and specificity for the control of CVL in scenarios of low or moderate transmission.

FOLLOW-UP QUESTION 4.2

The authors of this study report that the prevalence of canine *Leishmania* infection in endemic areas of Brazil shows wide variation from 1% to 67%. Which combination test strategy presents the greatest risk of culling (euthanatizing) uninfected dogs over this range? See Answer 4.2 at the end of this chapter.

4.3.3 Herd Retest

Herd retest is a modification of serial testing in which test-negative rather than test-positive animals are retested with the same test, usually at regular intervals. *The net effect is to ask the herd to prove that it is free of the condition being sought.*

Although test-negative animals are retested, herd retest does not increase the sensitivity of the testing strategy at the level of the individual because (1) the same test is used and (2) retesting occurs after a fairly long interval. Thus, a false negative may not be detected or, at best, would not be detected until sometime later. However, herd retest should increase the sensitivity of the testing strategy *at the herd level* because it increases the likelihood of detecting an agent on premises that, for whatever reason (sampling, incubating infection, reintroduction of the agent), eluded detection earlier. Herd retest forms the basis of test and removal programs designed to eradicate disease on any scale.

It is important to note that as the prevalence of disease in a population decreases, the specificity of the test procedure used to identify infected individuals becomes increasingly important. For example, application of a test that is only 80% specific in a herd that is free of infection will yield 20% test-positive individuals, all of which would be false positives. In a test and removal program, these animals would be needlessly sent to slaughter, much to the embarrassment of animal health officials if follow-up definitive testing is performed post-mortem.

4.3.4 Assumption of Independence of Multiple Test Results

Ideally, when multiple tests are used in parallel or in series, each test should measure a unique indicator of the health status of the individual. Body temperature, pulse, and respiratory rates are, by and large, independent measures. Immunologic tests also measure unique attributes of the individual when they depend on differences in the nature of the immune response, such as cellular versus humoral immunity, IgM versus IgG antibody, or differences in titer. If the assumption that the tests are completely independent is wrong, calculation of the probability of disease from several tests would tend to overestimate the value of the multiple testing strategy.

4.4 WORKING WITH DIFFERENTIAL LISTS

4.4.1 Rule-Ins and Rule-Outs: The Choice of Sensitive or Specific Tests

The choice of a particular diagnostic test, or the interpretation of a test result, should be made within the context of the clinical situation. A negative result on a highly sensitive test, i.e., one that is usually positive in the presence of disease, is frequently used to rule out a disease on a differential list during the early stages of a diagnostic workup. This concept was first introduced in Example 1.3, where the

absence of cutaneous lesions, present in more than three-quarters of dogs with leishmaniasis, could be used to lower the disease's ranking on the differential list. Tests of high sensitivity are also useful when there is an important penalty for failure to detect a particular disease, e.g., when the cost of a false negative result exceeds that of a false positive.

However, sensitivity should not be the sole criteria for choosing a test. As discussed earlier, use of a test of high sensitivity but low specificity in a disease eradication program may result in the removal from the herd of an unacceptable number of animals with false-positive results. A positive result on a highly specific test, i.e., one that is rarely positive in the absence of disease, is useful to confirm (or rule in) a diagnosis that has been suggested by other data. Tests of high specificity are especially useful when a false-positive diagnosis can result in physical, emotional, or financial loss to the patient or owner.

Two simple acronyms to facilitate remembering the above relationships are "**SnNout**" and "**SpPin**." In expanded form, "SnNout" states that negative results on tests of high sensitivity are best for ruling out a target disorder. "SpPin" states that positive results on tests of high specificity are best for ruling in a target disorder (Sackett et al., 1991).

When we simply wish to choose the test with the best overall performance (fewest diagnostic errors), test accuracy provides the best criteria. Even though the accuracy of a diagnostic test varies with the likelihood of the target disorder in the individual(s) being tested, the range of variability is less than positive or negative predictive values. Test accuracy provides a simple answer to the question: *"What proportion of test results are likely to be correct?"*

4.5 SCREENING FOR DISEASE

4.5.1 Definitions

When apparently healthy individuals or groups of individuals are systematically tested for the purpose of detecting certain characteristics or health problems, the process is referred to as **screening**. When screening tests are applied to large, unselected populations, this testing strategy is referred to as **mass screening or surveillance**. Mass screening is sometimes employed in state and federal disease control or eradication programs, such as the state USDA-APHIS national bovine brucellosis eradication program. APHIS conducts surveillance on animals at slaughter by testing their blood for the presence of *Brucella* antibodies. Around two million animals are tested annually. This testing strategy is more than sufficient to detect one infected animal out of 100,000 (or 0.001%) in the slaughter population at a confidence level of 95%, assuming a testing regime with 83% sensitivity. Estimated prevalence in the United States as of 2018 was 0.002%.

In other scenarios, only a "statistically representative" sample of the herd or flock is sampled with the objective of identifying affected populations rather than individuals. This strategy for herd testing was initially employed at the launch of the USDA's Aujeszky's Disease (Pseudorabies) Eradication Program in 1989. In 2004, all 50 states attained pseudorabies-free status in commercial production swine herds. However, sporadic infections continue to be found in outdoor production herds or swine that have access to the outdoors, especially where contact with feral swine is possible. Statistical sampling continues to be used in the current National Pseudorabies Surveillance Plan. Sampling of groups of individuals may also be accomplished by pooling of samples, as in testing of milk samples from bulk tanks for excessive bacterial counts or violative antibiotic residues.

Identification of an affected population may be followed with **case finding**. Case finding is a strategic form of screening targeted at individuals or groups suspected to be at high risk of infection or disease because of association with known infected or diseased individuals or groups, or through other forms of exposure. Case finding may be part of a traceback investigation of herds suspected of being sources of infected individuals involved in a disease outbreak. Case finding may also be employed during a foodborne disease outbreak investigation to identify as many affected individuals as possible.

The ongoing systematic and continuous collection, analysis, and interpretation of health data for the purpose of monitoring the spatial and temporal patterns of one or more diseases and their associated risk factors is referred to as **epidemiologic surveillance** (**surveillance**) or **monitoring**. An example is the systematic reporting and recording of cases of notifiable animal diseases by veterinary diagnostic laboratories. Surveillance data contribute to our understanding of the natural history of disease and are useful in the planning, implementation, and evaluation of disease control measures. Surveillance data provide the scientific basis for political, social, or economic decision-making.

EXAMPLE 4.3: HOW COMMON IS RABIES IN THE U.S. ANIMAL POPULATION?

Background: The rabies virus is one of the most successful zoonotic disease agents globally, with more than 30 reported animal reservoir species and near-global distribution (Ma et al., 2018). Human and animal rabies have been nationally notifiable conditions in the United States since 1944. Animal rabies surveillance is primarily a passive, laboratory-based system that comprises >130 state health, agriculture, and university pathology laboratories. These laboratories perform the standard direct fluorescent antibody test. In addition, as a component of a large-scale oral rabies vaccination program, the USDA Wildlife Services tests animals collected through enhanced (or active) rabies surveillance in selected geographic regions with the direct rapid immunohistochemical test. Enhanced rabies surveillance includes focused efforts to test abnormally behaving rabies vector species with no known human or domestic animal exposure as well as road-killed specimens, target species collected in a localized rabies focus area, and nuisance or hunter-collected animals. Enhanced surveillance serves as a complement to public health surveillance because efforts are concentrated on testing wildlife that would not likely be tested through exposure-based surveillance.

Objectives: Ma et al. (2018) summarized the results of rabies surveillance compiled by the U.S. Centers for Disease Control and Prevention (CDC) for 2016.

Study Design: Cross-sectional (prevalence survey based on surveillance data).

Methods: During 2016, a total of 95,424 animal samples were submitted for laboratory testing in the United States and territories, of which 93,535 (98.0%) were considered suitable for testing and included 6829 animals tested by the USDA Wildlife Services. The authors calculated percentages of rabid animals for each species based on the total numbers of animals tested, with only those animals with positive or negative test results included in the denominator. The results of rabies virus variant (RVV) typing were also reported when available.

Results: Table 4.5 compares test results for 2016 with data from 2011 through 2015. Wildlife accounted for 91.4% of rabies cases reported in 2016. Bats were the most frequently reported rabid animal in the United States, representing 33.5% of all animal rabies cases detected, followed by raccoons (28.6%), skunks (21.0%), foxes (6.4%), and domestic cats (5.2%). The percentage of rabid bats (6.9%) among the total submitted for testing was significantly higher than the mean percentage of 6.2% during the previous 5 years. Statistically significant declines in the percentage of test-positive samples were recorded for several other species (cattle, raccoons, skunks, foxes). Overall, the number of rabid animals and percentage of test

TABLE 4.5

Number of Animals Reported to be Rabid in the United States and Percentages of Samples Tested for Rabies that Yielded Positive Results for 2011 through 2016

	2016		2011–2015			
			No. of Rabid Animals		Percentage of Samples with Positive Results	
Animals	No. of Rabid Animals	Percentage of Samples with Positive Results	Mean	95% CI	Mean	95% CI
Domestic Animals						
Cats	257	1.2	265	243–286	1.1	1.1–1.2
Cattle	70	5.6[a]	86	70–102	6.9	5.9–7.8
Dogs	58[a]	0.3	74	63–85	0.3	0.3–0.4
Horses and donkeys	23	3.1	32	20–44	3.9	2.7–5.1
Sheep and goats	13[a]	2.3	10	8–12	2.0	1.4–2.6
Wildlife						
Raccoons	1403[a]	11.7[a]	1855	1727–1982	14.4	13.3–15.5
Bats	1646	6.9[a]	1624	1494–1753	6.2	5.9–6.5
Skunks	1031[a]	23.8[a]	1513	1420–1607	30.6	28.9–32.3
Foxes	313	17.2[a]	349	310–389	19.2	18.3–20.0
All rabid animals	4910[a]	5.3[a]	5920	5698–6142	5.9	5.7–6.2
Rabid domestic animals	423[a]	0.9	469	435–504	1.0	0.9–1.0
Rabid wildlife	4487[a]	9.6[a]	5450	5256–5645	10.9	10.6–11.2

Source: Ma X et al. *J Am Vet Med Assoc* 2018;252:945–957. With permission.

[a] Significantly ($p < 0.05$) different from mean value for 2011 through 2015.

positive animals, both domestic and wildlife, declined significantly in 2016 when compared with the mean for the previous 5 years.

Conclusions and Significance: The study confirmed that distinct RVVs continue to circulate among the major terrestrial animal reservoirs in the United States (raccoons, skunks, gray and arctic foxes, and mongooses [Puerto Rico]) in geographically definable regions (Table 4.6). In addition to the terrestrial RVVs, there are multiple RVVs associated with bats.

FOLLOW-UP QUESTION 4.3

Can the numbers and/or percentages calculated from the surveillance data summarized in Table 4.5 be interpreted as the incidence (new infections) of rabies in the animal populations from which the samples were drawn? (*Hint:* See Chapter 5, "Measuring the Commonness of Disease for a clue.") See Answer 4.3 at the end of this chapter.

TABLE 4.6
Rabies Virus Variants Identified in Domestic and Wild Animals in 2016

	Domestic Animals						Wildlife						
Variant	**Cats**	**Cattle**	**Dogs**	**Horses and Donkeys**	**Sheep and Goats**	**Other Domestic[a]**	**Raccoons**	**Bats**	**Skunks**	**Foxes**	**Other Wild[b]**	**Rodents and Lagomorphs[c]**	**Total**
Raccoon	49	18	15	2	2	2	352	0	153	110	8	5	716
South central skunk	24	9	11	2	1	0	17	0	275	17	2	0	358
North central skunk	2	5	9	1	0	0	0	0	34	1	0	0	52
California skunk	0	0	0	0	0	0	0	0	0	0	0	0	0
Arctic fox	0	0	0	0	0	0	0	0	0	8	2	0	10
Arizona gray fox	0	0	0	0	0	0	0	0	0	4	3	0	7
Texas gray fox	0	0	0	0	0	0	0	0	0	0	0	0	0
Bat	0	0	0	0	0	0	0	488	1	4	0	0	493
No variant reported	182	38	23	18	10	0	1034	1158	568	169	30	44	3274
Total infected	257	70	58	23	13	2	1403	1646	1031	313	45	49	4910
Variant typed (%)	29.2	45.7	60.3	21.7	23.1	100.0	26.3	29.6	44.9	46.0	33.3	10.2	33.3
Variant Typed (%), 2013–2015													
Mean	27.6	43.3	46.8	45.6	40.5	68.9	18.6	25.4	43.8	23.7	14.0	16.4	28.4
95% CI	22.8–32.5	31.8–54.9	37.3–56.3	36.2–55.0	23.1–57.9	41.1–96.7	17.2–20.0	21.8–29.0	41.0–46.5	18.5–28.8	10.5–17.4	10.1–22.6	27.1–29.7

Source: Ma X et al. *J Am Vet Med Assoc* 2018;252:945–957. With permission.

[a] Other domestic includes two alpacas with the raccoon RVV.

[b] Other wild includes two antelopes with the raccoon and south central skunk RVVs, two bobcats with the Arizona gray fox RVV, one bobcat with the raccoon RVV, one bobcat with the south central skunk RVV, one coyote with the Arizona gray fox RVV, four coyotes with the raccoon RVV, two deer with the raccoon RVV, and two wolves with the arctic fox RVV.

[c] Rodents and lagomorphs include two beavers with the raccoon RVV and three groundhogs with the raccoon RVV.

4.5.2 Test Criteria

Several criteria are used to evaluate the suitability of a diagnostic test for screening apparently normal populations. First, the test should be sensitive and specific. Because the prevalence of the condition being tested for will usually be low, the positive predictive value of the screening test will also be relatively low, regardless of its specificity. This effect can be diminished by restricting testing to high-risk groups. In addition to its performance characteristics, the test must be inexpensive, very safe, and acceptable to both clients and practitioners.

4.6 INCREASING THE PREDICTIVE VALUE OF DIAGNOSTIC TESTS

Considering the relationship between prevalence (pretest probability) of disease and the predictive value of a test, it is obviously to the clinician's advantage to apply diagnostic tests to patients with an increased likelihood of having the target disorder being sought. As a rule, tests are not ordered until the patient has undergone a thorough history and physical examination. Being a member of a high-risk group increases the positive predictive value of diagnostic tests. Consequently, clinicians should be aware of risk factors for specific diseases and the corresponding confirmatory diagnostic tests.

The referral process, such as that which contributes to the case load of veterinary teaching hospitals, increases the likelihood of finding significant disease. Consequently, more aggressive use of diagnostic tests might be justified in these settings versus the typical walk-in community practice. The same tests performed on a routine basis on all patients would have a lower predictive value because of the lower prevalence of disease.

> ***Being a member of a high-risk group increases the positive predictive value of diagnostic tests. Consequently, clinicians should be aware of risk factors for specific diseases and the corresponding confirmatory diagnostic tests.***

To illustrate the effect of pretest findings upon the predictive value of diagnostic tests, the reader is referred back to Table 4.2, where clinical staging is used to guide the interpretation of AGP levels in patients suspected of suffering from FIP. As the likelihood of FIP based on epidemiologic factors and clinical findings increases, the post-test probability of FIP for any AGP value increases dramatically. In contrast, for a patient with a very low pretest probability of FIP, the post-test probability does not increase noticeably even at the highest AGP value. Thus, predictive values and post-test probabilities of disease are more sensitive to variations in the prevalence (pretest probability) of disease than to the magnitude of a test result.

4.7 COMMUNICATION OF DIAGNOSTIC TEST RESULTS

Uncertainty is an inherent part of medical practice, and the interpretation and reporting of diagnostic test results is no exception. In most cases, there is a chance that the conclusion drawn from a test result is incorrect. It is important, therefore, to not only accurately communicate test results with colleagues and clients, but also to express the level of certainty of the associated diagnosis. Diagnostic certainty is usually conveyed using terms such as "likely," "probable," "consistent with," or "suggestive" (Christopher and Hotz, 2004). Although such terms are qualitative and imprecise expressions of probability, they may be preferable when people are uncomfortable with probabilities, or where numbers or percentages imply a level of precision that is unwarranted by the available information. Ultimately, whether words or numbers are used to express probability, they should communicate the intended information effectively and be clearly understood by and have value for the clinician or client.

EXAMPLE 4.4: HOW CAN THE REPORTING OF CLINICAL PATHOLOGY TEST RESULTS BE IMPROVED?

Background: Clinical pathologists use descriptive terms or modifiers to express the probability or likelihood of a cytologic diagnosis. Words are imprecise in meaning, however, and may be used and interpreted differently by pathologists, clinicians, and clients.

Objectives: Christopher and Hotz (2004) sought to (1) assess the frequency of use of 18 modifiers (terms) used to express diagnostic certainty; (2) determine the probability of a positive diagnosis implied by the modifiers; (3) identify preferred modifiers for different levels of probability; (4) ascertain the importance of factors that affect expression of diagnostic certainty; and (5) evaluate differences based on gender, employment, and experience.

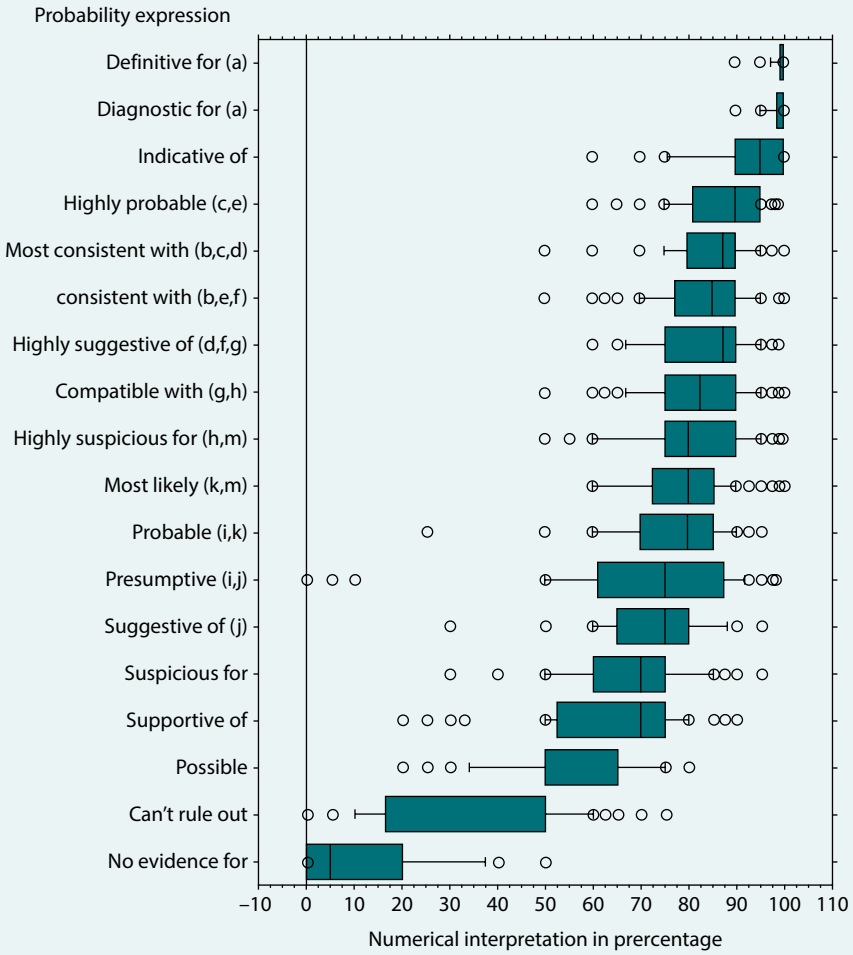

FIGURE 4.4 Numerical percentages attributed to 18 expressions of probability used to express uncertainty about a positive diagnosis. Boxes indicate the 25th to 75th percentile and median values; lines extend to the 10th and 90th percentile values; circles indicate outlying values. Expressions with the same letters in parentheses are not significantly different on the basis of ANOVA of transformed probabilities. (From Christopher MM and Hotz CS. *Vet Clin Pathol* 2004;33:84–95. With permission.)

TABLE 4.7

Summary of Preferred Modifiers for Expressing the Probability of a Positive Diagnosis at Defined Percentage Levels[a]

Modifier	100%	95%	75%	50%	25%	5%	0%
(No modifier)	24	3	0	0	0	0	2
Diagnostic for	37	3	0	0	0	0	0
Highly suggestive of	0	14	6	0	0	0	0
Most consistent with	1	13	4	2	0	0	0
Consistent with	5	14	9	2	0	0	0
Probable	0	12	23	4	1	0	0
Suggestive of	0	0	19	14	2	0	0
Possible	0	0	1	44	27	5	1
Can't rule out	0	0	0	6	21	30	2
Unlikely	0	0	0	0	6	10	0
No evidence for	0	0	0	0	1	11	42
Other	20	33	31	20	20	21	29
Do not express this level of probability	4	0	0	0	1	2	1
Number of different terms used	11	18	16	15	21	21	26
Total number of responses	91	92	93	92	78	77	76

Source: Christopher MM and Hotz CS. *Vet Clin Pathol* 2004;33:84–95. With permission.

[a] Shaded boxes indicate the responses of >50% of respondents for each numerical probability.

Study Design: Cross-sectional (survey).

Methods: The authors surveyed 202 clinical pathologists who were board-certified by the American College of Veterinary Pathologists (Clinical Pathology). Surveys were distributed in October 2001 and returned by e-mail, fax, or surface mail over a 2-month period. Results were analyzed by parametric and nonparametric tests.

Results: The survey response rate was 47.5% (n = 96) and primarily included clinical pathologists at veterinary schools (n = 58) and diagnostic laboratories (n = 31). There was considerable variability in the numerical percentage assigned to each of the 18 modifiers (Figure 4.4). Ninety of 96 (96.8%) respondents preferred words to numbers or percentages for expressing probability in cytology reports, and 10 terms expressing 7 probability levels ranging from 1% to 100% were preferred by ≥50% of respondents and recommended by the authors for expressing the probability of a positive diagnosis (Table 4.7).

Conclusions and Significance: The authors conclude that because of wide discrepancy in the implied likelihood of a diagnosis using words, defined terminology and controlled vocabulary may be useful in improving communication and the quality of data in cytology reporting. Vagueness inherent in the use of words to express diagnostic uncertainty is not necessarily bad, as it depends on the nature of the situation (as when tests are used as a screening test) and the quality of the information base.

FOLLOW-UP QUESTION 4.4

The conclusions of this study were based on a survey of veterinary clinical pathologists. What additional information might be useful in a follow-up study? See Answer 4.4 at the end of this chapter.

REFERENCES

Arruda MM, Figueiredo FB, Marcelino AP et al. Sensitivity and specificity of parallel or serial serological testing for detection of canine *Leishmania* infection. *Mem Inst Oswaldo Cruz*. 2016;111: 168–173. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4804499/

Christopher MM and Hotz CS. Cytologic diagnosis: Expression of probability by clinical pathologists. *Vet Clin Pathol* 2004;33:84–95.

Christopher MM, Hotz CS, Shelly SM et al. Use of cytology as a diagnostic method in veterinary practice and assessment of communication between veterinary practitioners and veterinary clinical pathologists. *J Am Vet Med Assoc* 2008;232:747–754.

Christopher MM, Hotz CS, Shelly, SM et al. Interpretation by clinicians of probability expressions in cytology reports and effect on clinical decision-making. *J Vet Intern Med* 2010;24:496–503. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1939-1676.2010.0492.x

de Arruda MM, Cardoso FA, Hiamamoto RM et al. Validity and reliability of enzyme immunoassays using *Leishmania major* or *L. infantum* antigens for the diagnosis of canine visceral leishmaniasis in Brazil. *PLOS ONE* 2013;8:e69988. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3724735/

Fagan TJ. Nomogram for Bayes's theorem. *N Engl J Med (letter)* 1975;293:257.

Ma X, Monroe BP, Cleaton, JM et al. Rabies surveillance in the United States during 2016. *J Am Vet Med Assoc* 2018;252:945–957.

Morris D. *Bayes' Theorem: A Visual Introduction for Beginners*. Blue Windmill Media; 2016.

Sackett DL, Haynes RB, Guyatt GH et al. *Clinical Epidemiology: A Basic Science for Clinical Medicine*, 2nd ed. Boston: Little, Brown and Co; 1991.

Saverio P, Alessia G, Vito T et al. Critical assessment of the diagnostic value of feline α_1-acid glycoprotein for feline infectious peritonitis using the likelihood ratios approach. *J Vet Diagn Invest* 2007;19:266–272. Full text available at: https://journals.sagepub.com/doi/full/10.1177/104063870701900306

Tasker S. Diagnosis of feline infectious peritonitis: Update on evidence supporting available tests. *J Feline Med Surg* 2018;20:228–243.

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 4.1: What's missing from Table 4.2 are the values for the likelihood ratios corresponding to each serum α_1-acid glycoprotein (AGP) measurement. These likelihood ratios were derived from their relationship with AGP concentrations depicted in Figure 4.2b and used to estimate post-test probabilities using either Bayes' formula or the nomogram depicted in Figure 4.1. Their exclusion does not diminish the value of Table 4.2, but the reader should be aware of how the post-test probabilities were derived.

Answer 4.2: Inappropriate culling of uninfected dogs is best avoided by choosing tests, or test combinations, that result in the highest positive predictive value (PPV); i.e., we can be most confident that a positive test result is correct. Based on the sensitivity analysis in Figure 4.3, parallel testing resulted in the lowest PPV of all combination test scenarios and is most likely to lead to the euthanasia of uninfected dogs.

Answer 4.3: Incidence is the proportion of at-risk individuals that develop a condition of interest over a defined period of time. Percentages reported in Table 4.5 cannot not be interpreted as the incidence (new infections) of rabies in these animal populations because (1) we do not know the size of the population from which samples were drawn and (2) most public health programs only test animals involved in potential exposure of humans or domestic animals to rabies. Therefore, the cases reported here may not represent the true number of animal rabies cases within the population at large.

Answer 4.4: It would be useful to know how veterinary practitioners interpret the terminology employed by clinical pathologists when reporting test results and how this might influence their treatment advice to clients. These questions were, in fact, answered in follow-up studies by the same

authors (Christopher et al., 2008, 2010). The most common reasons for practitioners to consult with a clinical pathologist were to discuss a discrepancy between clinical and cytologic findings, to clarify a diagnosis, and to ascertain the pathologist's confidence in a diagnosis. Ultimately, the terminology used would most likely influence a client's decision regarding treatment or euthanasia. The authors concluded that improving communication between veterinary practitioners and veterinary clinical pathologists through standardized terminology could enhance the diagnostic value of cytologic examinations, improve clinical decision-making, and enhance clinical outcomes.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

5 Measuring the Commonness of Disease

5.1 INTRODUCTION

Until now we have focused on the diagnosis of disease. We now turn our attention to measuring the frequency of disease events. Comparison of disease frequency in different groups forms the basis for assessing the risk of contracting a disease and its cause, prognosis, and response to treatment—the subjects of the next four chapters. Frequencies thus play a pivotal role in veterinary medical decision-making.

5.2 EXPRESSING THE FREQUENCY OF CLINICAL EVENTS

5.2.1 Proportions, Rates, and Ratios

The frequency of clinical events is usually expressed as a proportion, with cases as the numerator and **population at risk** as the denominator. These proportions are commonly referred to as **rates**, although the latter term is more appropriately reserved for those proportions that include a time component, e.g., that express the occurrence of new events in a population over a defined time interval. The reason for this distinction will be discussed further below when prevalence and incidence rates are compared.

A **rate** is not the same thing as a **ratio**. In the case of a rate, the numerator is included in the denominator, while in a ratio, the numerator and denominator are mutually exclusive. In other words.

$$\text{Rate} = \frac{a}{a+b} \quad \text{Ratio} = \frac{a}{b}$$

and in the special case for disease

$$\text{Rate} = \frac{\text{Affected}}{\text{Affected} + \text{Unaffected}} \quad \text{Ratio} = \frac{\text{Affected}}{\text{Unaffected}}$$

An example of a rate is the proportion of students enrolled in U.S. veterinary colleges that are male or female. An example of a ratio is the comparison of the frequency of male to female veterinary students, or vice versa. This chapter focuses on rates. Ratios will be used in the following chapter to estimate risks of clinical events.

EXAMPLE 5.1: WHAT IS THE GENDER DISTRIBUTION OF VETERINARY STUDENTS IN THE U.S.?

During the 1985–1986 academic year, the proportion (a **rate**) of female students (50.8%) enrolled in U.S. veterinary medical colleges surpassed that of males (49.2%) for the first time (AVMA, 1986; AAVMC, 2018). The corresponding **ratio** of female to male students was 1.033 to 1. The proportion of female students has steadily increased since then to approximately 80% for the 2017–2018 academic year at the 30 U.S. colleges of veterinary medicine (Figure 5.1).

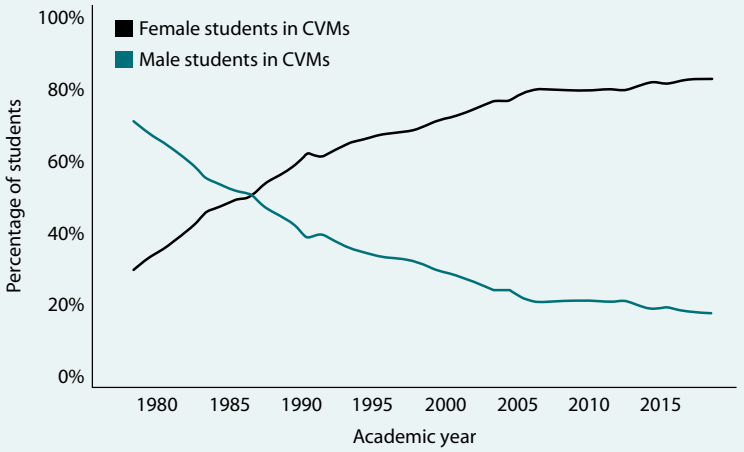

FIGURE 5.1 Gender representation at the 30 U.S. colleges of veterinary medicine 1978–2018. (From AAVMC. *Annual Data Report 2017–2018*. Washington, DC: Association of American Veterinary Medical Colleges; 2018. With permission.)

FOLLOW-UP QUESTION 5.1

What is your estimate of the ratio of female to male students for the 2017–2018 academic year? See Answer 5.1 at the end of this chapter.

A rate is not the same thing as a ratio. In the case of a rate, the numerator is included in the denominator, while in a ratio, the numerator and denominator are mutually exclusive.

Veterinarians regularly use a number of rates. Some are **vital statistics rates**, which provide indirect evidence of the health status of a population. Other rates may be classified as **morbidity rates**, i.e., direct measures of the commonness of disease. Among the latter, the three most commonly used are prevalence, incidence, and attack rate. Several of the more commonly used vital statistics and morbidity rates are listed in Table 5.1.

5.2.2 Prevalence, Incidence, and Attack Rate

Prevalence is the proportion of sampled individuals that possesses a condition of interest at a given point in time. It is measured by a single examination of each individual of the group. Prevalence is a static measure in which the time unit is short (one day or a few days). It can be likened to a "snapshot" of the population and includes both old and new cases. It is a measure of the likelihood of being a case at a given point in time.

In some cases, a distinction is made between animal level and herd prevalence. **Animal level prevalence** expresses the proportion of animals (the population at risk) that possess the condition of interest. When these animals reside within the same herd, then this is referred to as **within-herd prevalence**. In contrast, **between-herd prevalence** expresses the proportion of all herds in which one or more animals possess the condition of interest. Within-herd prevalence is useful when describing the spatial distribution of disease. As factors determining the level of within-herd prevalence are usually distinct from those determining between-herd prevalence, there may be a marked discrepancy between the two figures in any given region. See Example 5.6 later in this chapter to better understand the difference among these three methods of expressing prevalence.

TABLE 5.1
Commonly Used Vital Statistics and Morbidity Rates in Veterinary Medicine

Rate and Its Calculation	Remarks
Vital Statistics	
Crude live birth rate: $\frac{\text{No. of live births}}{\text{Average population}} \times 10^x$	Useful as a measure of population increment due to natural causes
General fertility rate: $\frac{\text{No. of live births}}{\text{Average no. of females of reproductive age}} \times 10^x$	Frequently used as an index of overall herd reproductive performance
Crude death rate: $\frac{\text{No. of deaths}}{\text{Average population}} \times 10^x$	Useful as a measure of population loss due to natural causes
Morbidity/Mortality Rates	
Attack rate: $\frac{\text{No. of affected individuals during an outbreak}}{\text{Population at risk at beginning of the outbreak}} \times 10^x$	Useful for identifying risk factors for a specific disease; restricted to outbreak investigation
Incidence rate: $\frac{\text{No. of new cases of a disease over a time interval}}{\text{Average population at risk during the time interval}} \times 10^x$	A dynamic measure of risk of acquiring disease over a given period; useful for monitoring the course of an epidemic; used in cohort studies to measure effect of suspected or known risk factors
Prevalence: $\frac{\text{No. of existing cases of a disease at a point in time}}{\text{Population at risk at the same point in time}} \times 10^x$	A static measure of the risk of having a particular disease at a given point in time; used in case control studies to measure effect of suspected or known risk factors
Case fatality rate: $\frac{\text{No. of deaths from a specified cause}}{\text{Total no. of cases of the same disease}} \times 10^x$	Useful for determining prognosis for a specific disease

Source: Adapted from Armstead WW. In Schwabe CW (ed). *Veterinary Medicine and Human Health*, 3rd ed. Baltimore: Williams & Wilkins; 1984.

Note: Although any time period could be used, for convenience all indices refer to a defined population of animals observed for 1 year unless otherwise stated.

Although prevalence is a snapshot of the disease status of a population, a series of prevalence measurements can be combined to obtain a picture of the occurrence of disease over time. This approach is particularly useful for detecting disease trends.

EXAMPLE 5.2: IS PASSIVE SURVEILLANCE AN EFFECTIVE STRATEGY FOR BOVINE BRUCELLOSIS ERADICATION?

Background: Bovine brucellosis, a zoonotic disease caused by *Brucella abortus*, is endemic in Colombia and many other South American countries (Cárdenas et al., 2018). An official control program has been in place in Colombia for more than 20 years, and vaccination of female cattle and buffalo between the ages of 3 and 8 months of age is compulsory. Surveillance consists of a passive reporting system to identify affected herds and a voluntary test-and-removal program with a goal of eradicating the infection from affected herds, with yearly recertification. Animals cannot be moved between farms unless they test negative for *B. abortus* infection.

Objectives: Cárdenas et al. (2018) evaluated the effectiveness of the Colombian *B. abortus* surveillance and control program over the 7-year period from 2006 to 2012.

Study Design: Cross-sectional (prevalence survey).

Methods: The investigators used data provided by the official veterinary service of Colombia (Instituto Colombiano Agropecuario, ICA). All samples were analyzed according to the World Organization for Animal Health manual of procedures for serological screening based on the rose Bengal plate test and indirect ELISA in series and complemented with competitive ELISA.

Results: The proportion of farms and cattle infected with *B. abortus* per region and per year are depicted in Figures 5.2a and 5.2b, respectively. Both graphs show some fluctuations over the study period but no clear trend. The percentage of positive farms averaged 22% in 2006 and 23% in 2012, and the percentage of positive animals averaged 4.7% in 2006 and 4.6% in 2012, indicating that the brucellosis control program had a low impact in Colombia over this time period. The level of infection and surveillance was found to vary among the different geographic regions, largely as a function of the intensity and nature (dairy versus beef) of livestock production.

Conclusions and Significance: The authors conclude that the current brucellosis control program, based on passive surveillance, has had little impact on the prevalence of *B. abortus* infection of cattle in Colombia over the 7-year observation period. They encourage a risk-based approach to brucellosis surveillance coupled with continuous evaluation of the effectiveness of control measures, training of laboratory and field personnel, and producer education.

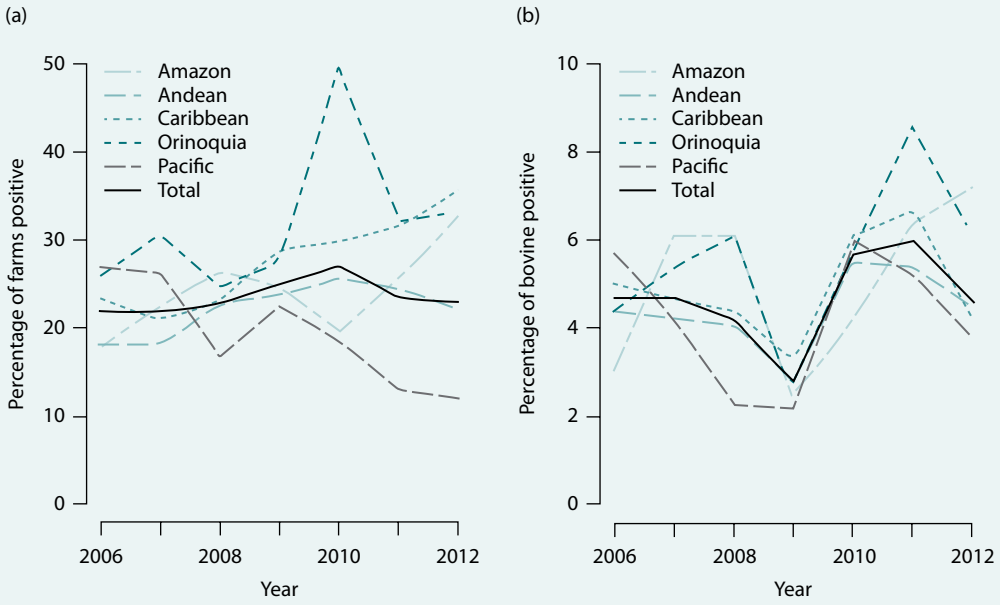

FIGURE 5.2 Proportion of farms (a) and cattle (b) that tested positive for *Brucella abortus* infection per region and year in Colombia. (From Cárdenas L et al. *Trop Anim Health Prod* 2018; 50: 19–27. With permission.)

FOLLOW-UP QUESTION 5.2

The proportion of farms and cattle infected with *Brucella abortus* depicted in Figures 5.2a and b are described by the authors as prevalence figures. As testing was conducted over a 7-year period, why aren't these proportions considered measures of incidence? (*Hint:* See the discussion of incidence in the next section for a clue.) See Answer 5.2 at the end of this chapter.

Prevalence represents the likelihood of being a case, whereas incidence represents the likelihood of becoming a case.

Incidence is the proportion of individuals in a susceptible population that develop a condition of interest over a defined period of time. Although birth rates, death rates, and similar vital statistics are based on new events, incidence is commonly understood to refer to disease events. Incidence takes into account new cases only, i.e., cases that have their onset during the specified time period. It is, therefore, a measure of the risk of becoming a case over a defined time period. An example would be the incidence of postoperative surgical site infections for different kinds of surgeries performed in a veterinary teaching hospital. The population at risk would be animals undergoing surgery during a specified time period. Surgeries might be further classified based on site, duration, etc. for comparative purposes.

Ideally, the population at risk is a cohort of all susceptible individuals at the beginning of the follow-up period. A **cohort** is a group of individuals who have something in common when they are first assembled, and who are then observed for a period of time to see what happens to them. Often, because of the difficulty of conserving the original composition of a cohort over follow-up periods of long duration, the denominator is expressed as the average population at risk during that period.

EXAMPLE 5.3: HOW COMMON IS POSTOPERATIVE REGURGITATION AND VOMITING IN DOGS AND HOW CAN IT BE PREVENTED?

Background: Postoperative regurgitation and vomiting (PORV) can lead to multiple adverse outcomes in canine patients, including aspiration pneumonia, esophagitis, esophageal strictures, esophageal perforation, and increased tension on suture lines, potentially resulting in prolonged hospitalization and increased treatment costs (Davies et al., 2015). The human counterpart is known as postoperative nausea and vomiting (PONV).

Objectives: Davies et al. (2015) conducted a study to determine the overall incidence of POVR in dogs and identify relevant risk factors.

Study Design: Historical cohort study.

Methods: Surgery reports for two cohorts of canine patients who had undergone nonelective neurologic, orthopedic, or soft tissue surgery over two 3-month periods (January 1 to March 31, 2000, and January 1 to March 31, 2012) at the Veterinary Teaching Hospital, Washington State University, were retrieved. Incomplete or missing medical records were excluded, as were reports for patients that had undergone elective procedures (i.e., ovariohysterectomy and neutering). A total of 111 patients in 2000 and 133 in 2012 met the inclusion criteria. The case definition for PORV used by the authors was any patient with vomiting or regurgitation within 24 hours after recovery from anesthesia. As no significant difference in the incidence of PORV or risk factor distribution between the two cohorts was detected, data from all 244 patients

were pooled for risk factor analysis. The data were subjected to univariate and multivariate analysis to identify potential risk factors for PORV.

Results: The overall incidence of PORV among the 244 patients was 12.3%, with 5.7% vomiting, 5.7% regurgitating, and 0.8% both vomiting and regurgitating. Univariate logistic regression identified the most significant risk factors to be gastrointestinal surgery; premedication without strong sedatives, including either an α_2-adrenoceptor agonist or American Society of Anesthesiologists (ASA) score of 4; history of vomiting or regurgitation; emergency surgery; neurologic surgery; sevoflurane inhalation anesthesia; and being sexually intact. Multivariate analysis, intended to control for interactions (such as age, breed, and weight) among variables, was not clinically useful owing to the low sensitivity and specificity of the model.

Conclusions and Significance: The authors recommend that antiemetic prophylaxis be considered in dogs undergoing gastrointestinal surgery and in those in which other risk factors are present.

FOLLOW-UP QUESTION 5.3

The study design employed by these investigators was prospective in that patients were enlisted into the study at the time of surgery and followed into their future. However, as it relied upon preexisting data recorded in the medical records years earlier, it varied in its level of detail. What would be an appropriate follow-up study design? (*Hint:* See Tables 1.4 and 1.5 for a guide to study designs employed in clinical research.) See Answer 5.3 at the end of this chapter.

The preceding example illustrates the value of detailed medical records for epidemiological research. Historical cohorts (see Chapter 6, "Risk Assessment and Prevention") of canine patients that had undergone non-elective surgery years earlier were assembled from historical records along with accurate follow-up data, enabling the investigators to explore the contribution of a variety of risk factors for POVR. This study is another example of the One Health concept. Post-surgical vomiting and regurgitation is of concern in both human and veterinary medicine, and the lessons learned are often transferable between disciplines.

Attack rate is a term that is often used to specify the proportion of a defined population affected during an outbreak. In its simplest form, the attack rate is equal to the total number of cases during the outbreak period divided by the number of individuals initially exposed, i.e., those present at the beginning of the outbreak. Since the attack rate is based only on new cases of the disease, it is comparable to incidence. Attack rate tables are particularly useful for evaluating the relative contribution of various risk factors to the onset or course of an epidemic or a food-borne disease outbreak. Measurement of risk is discussed in greater detail in Chapter 6.

5.3 MEASURING THE FREQUENCY OF CLINICAL EVENTS

5.3.1 Prevalence

Prevalence is measured by surveying a population, some of whose members are diseased and the remainder healthy, at a particular point in time. The proportion that are diseased constitutes the prevalence of the disease. Such "snapshots" of the population are referred to as **cross-sectional studies**.

Prevalence can be estimated through examination of a group of animals at a single point in time (**point prevalence**), a single examination of each of a series of animals seen over a period of time (**period prevalence**), or a combination of the two.

EXAMPLE 5.4: HOW COMMON IS ENTEROHEMORRHAGIC *ESCHERICHIA COLI* (EHEC) CONTAMINATION OF BEEF HIDES?

Background: Enterohemorrhagic *Escherichia coli* are human foodborne pathogens. Cattle are important reservoirs, and their hides can be contaminated by EHEC through direct contact with feces and contact with other cattle or animals, dust, or insects. Carcasses may become contaminated during the hide removal process at the processing plant.

Objectives: Schneider et al. (2018) sought to estimate the seasonal and regional occurrence of hide contamination by each of seven EHEC serogroups among U.S. market beef cows at slaughter.

Study Design: Cross-sectional (prevalence survey).

Methods: One hundred hides were sampled during each of four seasonal sampling events at two geographically distinct beef packing plants over a 10-month period. The source of cattle was categorized as either the southern (Plant A) or northern (Plant B) plains regions of the United States.

Results: At least one EHEC serogroup was detected on 630 of the 800 total hides sampled, yielding a **period prevalence** of 79% over the course of the study. Of the 400 samples each obtained at Plants A and B, 370 (93%) and 260 (65%) were positive for at least one EHEC serogroup, respectively. Seasonal contamination by individual serogroups ranged from to 0% to 66.5%, and 6% to 57.5% by geographic region. The relationship between season and region was different for each EHEC serogroup, suggesting that conditions that differ over time and place affect the risk that cattle will carry each EHEC serogroup.

Conclusions and Significance: This example illustrates how both period and point prevalence can be estimated from the same study. Even though sampling occurred over time (**period prevalence**), the results for each sampling event ("plant visit") represent the proportion of contaminated hides at a single point in time, i.e., **point prevalence**, thus permitting the investigators to make seasonal and geographic comparisons. The authors caution, however, that results were obtained by sampling cattle on only 8 days in four seasons in a single year. Specific results could have differed depending on the day of sampling.

FOLLOW-UP QUESTION 5.4

For this study, sampling occurred at two geographically distinct packing plants, intended to represent cattle arriving from northern and southern portions of the United States. The authors found a significant difference in the period prevalence of EHEC between the two packing plants. Besides the origin of the cattle, what other factors could have influenced this finding? See Answer 5.4 at the end of this chapter.

Prevalence studies can be based on the examination of a group of animals at a single point in time, on a single examination of each of a series of animals seen over a period of time, or a combination of the two.

5.3.2 Incidence

Incidence is measured by recording the occurrences of a condition of interest over time in a population initially free of the condition. This study design, called a **longitudinal or prospective study,** is discussed in detail in Chapter 6. Whereas time is assumed to be instantaneous in cross-sectional studies, it is a key component in the measurement and expression of incidence.

Incidence is commonly measured in one of two ways. In the first, a **defined group** of susceptible individuals, known as a **cohort**, is followed over time and each occurrence of the event of interest is recorded as it occurs. This approach is frequently used to determine the prognosis, with or without treatment, for a group of individuals known to be affected with a particular disease. Examples of this kind of study appear in Example 1.1 describing the incidence of anesthetic death (death rate) among dogs and cats during or before full recovery from anesthesia, and Example 5.3 describing the incidence of postoperative regurgitation and vomiting in dogs. In these examples, subjects were recruited into their respective studies at the time surgery was performed and were then followed for a defined period of time.

Incidence can also be measured by recording the number of new events occurring in an ever-changing population whose members are at risk for varying periods of time. This approach is useful for determining the effect of a risk factor on the subsequent incidence of disease in a dynamic population. In this case, the denominator of the incidence rate must be adjusted to account for the variable period of time that each animal is exposed to the risk factor. Sometimes the average number of animals present over the specified time interval is used as the denominator. A more accurate approach is to use **animal time at risk** rather than number of animals in the denominator. The resulting incidence rate is then referred to as an **incidence density**, and reflects the number of new events, or cases, per total number of animal days, weeks, months, or years at risk.

EXAMPLE 5.5: HOW EFFECTIVE IS NATURAL EXPOSURE FOR ESTABLISHING HERD IMMUNITY TO *ANAPLASMA MARGINALE*?

Background: Anaplasmosis is an infectious blood disease of cattle caused by the intraerythrocytic rickettsia *Anaplasma marginale*. The causative agent is transmitted biologically by ticks and mechanically by biting flies and veterinary or animal husbandry procedures (vaccination, blood sampling, dehorning, ear tagging, etc). The severity of disease is dependent upon a number of factors, including breed of cattle, strain virulence, and age at time of infection. Mature cattle may develop severe disease with high mortality, whereas younger cattle show minimal clinical signs. Recovered cattle develop a lifelong carrier state and are immune to subsequent clinical anaplasmosis. Bovine anaplasmosis is endemic in many parts of California where carrier cattle and various species of deer serve as reservoirs of infection for grazing cattle. Beef cattle managers often intentionally move herds of younger cattle to high risk, tick-infested pastures to encourage infection while they are still resistant to severe clinical disease.

Objectives: Following reports of an increased incidence of clinical anaplasmosis in California beef herds, Tucker et al. (2016) conducted a prospective cohort study to estimate the incidence of *A. marginale* infection, based on seroconversion, in heifers and mature cattle in a northern California beef cattle herd naturally exposed on tick-infested pastures. The authors hypothesized that the incidence density rate (IDR) for all age classes would be low, suggesting that a high proportion of seronegative cows would be at risk of clinical anaplasmosis.

Study Design: Cohort study.

Methods: A total of 143 Black Angus cattle (106 prebreeding heifers and 37 cows) were enrolled in one of three cohorts at various times during the course of the study. Cattle ages and pasture locations varied among the cohorts (Table 5.2). Serum samples were collected to determine *A. marginale* seroprevalence using a commercially available competitive enzyme-linked immunosorbent assay test kit. Repeat sampling of seronegative cattle was performed to determine the IDR from March through September 2013.

Results: The estimated incidence density rate of *A. marginale* infection for all groups, calculated as 1000 × (incident cases/cow-days at risk), was 8.17 (95% confidence interval: 6.04,

TABLE 5.2
Summary of Population at Risk, Incident Cases, Time at Risk, and Incidence Density Rate (IDR) of *Anaplasma marginale* Infection for Cattle Followed up to Determine Rate of *A. marginale* Infection

Age Group	Group 1 Heifers	Group 2 Heifers[a]		Cows
Pasture	C	A	B	C
Population at risk[b]	31	37	10	5
Incident cases	18	26	3	2
Days at risk	80	70	123	80
Incidence density rate	7.26	10.04	2.44	5.00

Source: Tucker TR et al. *Vet Med Int* 2016; 6186078. With permission.

[a] Group 2 heifers were grazed on pasture A for the first 70 study days and on pasture B for the remaining 123 study days.

[b] Number of cattle seronegative at start of follow-up period.

10.81) cases per 1000 cow-days during the study period. Study cattle became *A. marginale* seropositive and likely carriers protected from severe clinical disease that might have occurred had they been first infected as mature adults. Results for individual cohorts varied and are summarized in Table 5.2.

Conclusions and Significance: No clinical cases of anaplasmosis were observed during the course of this study, suggesting that the IDR measured through natural exposure was sufficient to establish herd immunity and that no changes to the herd management protocols were warranted. See Chapter 12 for a discussion of herd immunity and its role in preventing disease spread within a population.

FOLLOW-UP QUESTION 5.5

The authors propose that a "benchmark" *A. marginale* seroprevalence of 60% during the first year of age would be sufficient to prevent anaplasmosis outbreaks in a herd. The mean overall IDR in this study was estimated to be 8.17 cases per 1000 cow-days. Assuming a minimum tick-borne transmission period of 90 days, what proportion of the herd would be infected by 1 year of age? See Answer 5.5 at the end of this chapter.

5.4 FACTORS AFFECTING THE INTERPRETATION OF INCIDENCE AND PREVALENCE

5.4.1 Temporal Sequence

Prevalence studies can be used to obtain a static picture of a situation at a fixed point in time, i.e., a snapshot of the population. Examples are provided in Chapters 3 and 4 in which prevalence data were used to evaluate the performance of diagnostic tests. Other examples are the routine surveillance activities of animal disease control programs, diagnostic laboratories, and veterinary teaching hospitals.

Prevalence studies can also be used to examine the possible causal relationship (association) between suspected risk factors and the health status of a population. Unlike incidence studies, this relationship was not studied over time. Thus, we can only infer which came first, the putative cause or the outcome of interest. These relationships are further depicted in Figure 5.3.

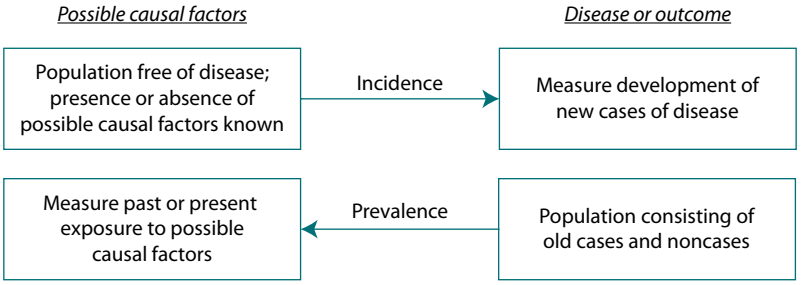

FIGURE 5.3 Temporal relationship between possible causal factors and disease: approaches based on incidence versus prevalence.

An important limitation of prevalence studies of cause is that one must infer the sequence of events.

5.4.2 Disease Duration

The population included in the numerator of an incidence rate may differ from that in a prevalence rate. In an incidence study, new cases are recorded as they occur over time. In a prevalence study, it is difficult to distinguish new from old cases. Furthermore, if a disease is of short duration or fatal, some cases may be missed because they are no longer detectable at the time the prevalence study is conducted.

The prevalence of a disease in a population may be higher or lower than incidence, depending on the average duration of the disease. Diseases that are rapidly fatal, such as rabies (see Example 4.3), or of short duration, such as bovine mastitis, might have a higher incidence than prevalence. Chronic diseases, such as bovine anaplasmosis (see Example 5.5) and other parasitic infections, might be readily detected for long periods of time, and would be more likely to appear in a prevalence study.

The prevalence of a disease in a population may be higher or lower than incidence, depending on the average duration of the disease.

5.4.3 Relationship among Incidence, Prevalence, and Duration of Disease

Since prevalence is the likelihood of being a case at any particular time, anything that increases the duration of disease will increase prevalence. Stated mathematically, prevalence can be estimated by multiplying incidence times the average duration of disease (prevalence = incidence $\times$ average duration of disease). The equation can be rearranged to calculate any one of the parameters of interest.

5.4.4 True versus Apparent Prevalence

Chapters 3 and 4 discussed how results derived from tests of less than 100% sensitivity and specificity may not indicate the **true prevalence** of disease. These tests measure the **apparent prevalence** of disease in a population, as distinguished from true prevalence, which is usually estimated through the use of an appropriate gold standard. If estimates of the sensitivity and specificity of a diagnostic test are available, then true prevalence can be estimated from apparent prevalence. This estimate would be useful during the course of a disease eradication program, where the actual level of disease still

present in the test population must be known as accurately as possible. The Rogan-Gladen estimator (Rogan and Gladen, 1978) can be used to estimate true prevalence from apparent prevalence as follows:

$$\text{True Prevalence} = \frac{\text{Apparent prevalence} + \text{Specificity} - 100\%}{\text{Sensitivity} + \text{Specificity} - 100\%}$$

This equation will not tell us whether a given test result is correct, but it does provide a better estimate of the true prevalence of disease.

EXAMPLE 5.6: IS THE DISCREPANCY BETWEEN APPARENT AND TRUE PREVALENCE IMPORTANT?

Background: The Kars region of Turkey is a major sheep production area. A significant proportion of the sheep of the region are afflicted with Johne's Disease (paratuberculosis), caused by *Mycobacterium avium* subsp. *paratuberculosis* (*Map*), a widespread zoonotic disease of livestock, particularly sheep and cattle. The actual prevalence of infection is not known, but it is an important measure of the both magnitude of the problem and the success of any control program.

Objectives: Buyuk et al. (2014) conducted a study to determine the prevalence of infection within herds, among herds, and among animals overall using a commercial enzyme-linked immunosorbent assay kit.

Study Design: Cross-sectional (prevalence survey).

Methods: A cross-sectional sample of 450 blood samples from sheep 24 months of age or older randomly selected from within 26 herds was tested. The reported sensitivity and specificity of the ELISA test kit were 64% and 99%, respectively. Herds were declared positive if one or more sheep in the herd tested positive for *Map* antibodies. Fifteen herds with a total of 275 sheep had at least one *Map* positive reactor. The animal, within-herd, and between-herd apparent prevalences were calculated by dividing the number of test positive outcomes by the corresponding denominator, i.e., total number of sheep tested from all herds (n = 450), total number of sheep tested within the 15 positive herds (n = 275), and total number of herds tested (n = 26), respectively. True prevalence was estimated from apparent prevalence results using the Rogan-Gladen estimator (see above).

Results: Results are summarized in Table 5.3. A sample estimation of the true prevalence of *Map* infection, using test properties reported above and the apparent prevalence value for all animals from Table 5.3, follows:

$$\text{True Prevalence of } \textit{Map} \text{ infection in all sheep} = \frac{6.2\% + 99\% - 100\%}{64\% + 99\% - 100\%} = 8.3\%$$

The within-herd true prevalence for each of the 26 infected herds ranged from 3.7% to 40.7%, while between-herd true prevalence rose to 90% from an apparent prevalence estimate of 57.7%.

Conclusions and Significance: The authors concluded that *Map* infection is widespread in sheep herds in the Kars region of Turkey, justifying the need for an efficient control program.

TABLE 5.3
Apparent and True Animal, Within-Herd, and Between-Herd Prevalence Estimates for *Mycobacterium avium* subsp. *paratuberculosis* (*Map*) Infection of Sheep in the Kars Region of Turkey

Prevalence Type	Number Tested	Number Positive for Map	Apparent Prevalence		True Prevalence	
			Estimate (%)	95% CI	Estimate (%)	95% CI
Animal	450	28	6.2	4.3–8.8	8.3	4.7–11.8
Within-herd	275	28	10.2	7.1–14.3	14.6	8.9–20.2
Between-herd	26	15	57.7	38.9–74.5	90	59.8–120.1

Source: Buyuk F et al. *Veterinarni Medicina* 2014; 59: 331–335. With permission.

FOLLOW-UP QUESTION 5.6

All of the estimates of true prevalence in Table 5.3 are greater than the estimates for apparent prevalence based on the ELISA test kit results. What variable in the Rogan-Gladen estimator may account for this discrepancy? See Answer 5.6 at the end of this chapter.

5.4.5 Case Definition

In many instances, it is difficult to define a set of disease signs, referred to as the **case definition**, that will include all true cases of the disease and exclude similar, but unrelated, conditions. For example, in Chapter 3 (Table 3.2), a list of clinical and clinicopathologic findings associated with elapid snake envenomation in horses was presented. The percentage of horses exhibiting any one finding ranged from 10% to 96%. As the number of signs required to diagnose elapid snake envenomation increases, the definition becomes more and more restrictive and includes a progressively smaller number of cases.

5.4.6 Dangling Numerators

Expressing the frequency of disease as a rate or proportion, using appropriate denominators rather than in terms of absolute numbers, i.e., **dangling numerators**, permits comparisons of disease frequency in comparable populations. Comparing numbers of cases (numerator data) without taking into consideration the population at risk (denominator data) does not tell us anything about the risk of becoming (incidence) or of being a case (prevalence). For example, Table 4.5 in the previous chapter lists the number and percentage of rabid animals detected among domestic and wildlife species. Among domestic species, the number of rabid cats (numerator data) was far greater than that of any of the other domestic species, suggesting that cats were especially vulnerable to rabies virus infection. However, when interpreted in light of the number of samples submitted (denominator data), the percentage of positive feline samples was far below that for cattle and all but one of the other species/species groups tested.

5.4.7 Population at Risk

Incidence rates and prevalence must be interpreted in the context of the population at risk. If the population at risk differs significantly from one's own patients, then extrapolations may be meaningless. For example, because of the frequency of referral cases and usage patterns of veterinary services, the population of animals presented to the typical veterinary teaching hospital (VTH) is

not representative of the population as a whole. This does not mean that the VTH patient population cannot serve as a denominator. If we wish to know the frequency with which findings occur among individuals with particular diseases (sensitivity data), then patients must be the denominator. On the other hand, if we wish to know the prevalence of the condition in the general population, then we would have to change our sampling strategy.

It is seldom feasible to sample the entire population at risk. Typically, a representative sample is selected by a random procedure in which all individuals have an equal chance of being included in the study. Sampling techniques and statistics are discussed further in Chapter 9. Comparison of disease rates among different groups is fundamental to determining the presence, cause, source, or probable mode of transmission of a disease. When comparing rates, care should be taken to ensure that populations used as denominators are truly comparable.

Comparing numbers of cases without taking into consideration the population at risk does not tell us anything about the risk of becoming (incidence) or of being a case (prevalence).

5.4.8 Crude versus Adjusted Rates

Rates such as incidence, prevalence, and attack rate are considered **crude rates** when they are expressed in the standard format:

$$\frac{\text{Total number of affected individuals}}{\text{Total population}} \times \text{Multiplier}$$

It should be recognized that the crude rate summarizes the effects of two factors:

1. *Subgroup-specific rate*: The probability of the event occurring in each subgroup (or stratum) of a population (such as subgroups based on age, breed or sex), and
2. *Subgroup distribution*: The characteristics or distribution of the subgroups in the population under consideration.

Because a crude rate is a composite figure, it is necessary to disentangle these two factors before meaningful comparisons can be made between population groups. **Adjusted (standardized) rates** compensate for subgroup effects by converting their distribution to that of a **standard population**.

Age is one of the most important characteristics governing the distribution of health-related events. Before morbidity or mortality rates in two populations can be compared, account must be taken of differences in age. Consider the data in Table 5.4. A paradox is seen. Age-specific death rates were higher for calves in both age groups on Farm A, where antibiotics were given. Yet, the overall death rate was higher on Farm B, where antibiotics were not given to calves. The apparent advantage of antibiotic use in calves is the result of the difference in age distribution of calves in the two comparison groups (Farms A and B). As a matter of record, the original findings showed that overall mortality for live births was 7.6% among calves given antibiotics versus 5.2% among those not given antibiotics, i.e., antibiotics were being used therapeutically rather than prophylactically. Mortality figures were based on cohorts of calves followed from birth through 60 days of age (Oxender et al., 1973).

The effect of differences in age distribution among subgroups of calves in the preceding example is an example of **confounding**. In this case age is referred to as a **confounding factor** because it confounds or blurs the comparison of the two farms. When differences in the distribution of one or more host characteristics, such as age, occur among the groups we wish to compare, **adjusted (standardized) rates** should be used.

TABLE 5.4
Death Rates of Calves by Age on Two Farms According to Antibiotic Use

	Farm A		Farm B	
	Antibiotics Given to Calves[a]		Antibiotics Not Given to Calves	
Age Group	Population at Risk	Death Rate	Population at Risk	Death Rate
0–14 days	105	10.5	118	7.6
15–60 days	307	4.2	40	2.5
All ages	412	5.8	158	6.3

Source of age-specific death rates: Oxender WD et al. *J Am Vet Med Assoc* 1973;162:458–460.
[a] Antibiotics were being used therapeutically rather than prophylactically.

Because a crude disease rate is a composite figure reflecting two factors, namely specific disease rates and population compositions, it is necessary to disentangle the two factors before meaningful comparisons can be made between population groups.

5.5 ADJUSTED RATES: THE DIRECT METHOD

One method that can be used to adjust rates is referred to as the **direct method** (Fletcher et al., 2014; Gordis, 2014). To understand what is meant by an adjusted rate, it must first be recognized that a crude rate may be expressed as the weighted sum of **subgroup-specific rates**. Each component of the sum (crude rate) has the following form:

$$(\text{Proportion of the population in each subgroup}) \times (\text{Subgroup-specific rate})$$

The basic idea in computing direct rates for comparison of populations is to compute what the hypothetical crude rates would be for the populations if the confounding factor were similarly distributed among their respective subgroups. In other words, we force a comparison of populations based on a *common* distribution of the confounding factor.

To compute directly adjusted rates, we need only two basic pieces of information: (1) the subgroup-specific rates for each subgroup and (2) a standard population. The **standard population** is that common distribution whose primary purpose is to serve as a reference group or substitute for the different distributions of the populations being compared.

5.5.1 Age-Adjusted Rates

Age is one of the most common confounding factors that is adjusted for. In the following example, we calculate and compare age-adjusted rates using the data on calf mortality from Table 5.4. We arbitrarily define the standard population to be the sum of calves from the two farms in each age group. In this example, the method for calculating age-adjusted death rates involves three steps and is presented in Table 5.5:

1. Estimate the number of expected deaths for each age group by multiplying the standard population at risk by the observed death rate for each age-specific group.
2. Estimate the total number of expected deaths by adding expected deaths over all age-specific groups.
3. Estimate the direct rate by dividing the total expected deaths by the total standard population.

TABLE 5.5
Direct Adjustment of Death Rates among Calves on Two Farms According to Antibiotic Use

		Farm A		Farm B	
		Antibiotics Given to Calves		Antibiotics Not Given to Calves	
Age Group	Standard Population at Risk	Death Rate per 100	Expected Deaths	Death Rate per 100	Expected Deaths
0–14 days	223	10.5	23.4	7.6	16.9
15–60 days	347	4.2	14.6	2.5	8.7
Totals	570		38		25.6
Direct rate (per 100) for Farm A	$\frac{38}{570}$	= 6.7	Direct rate (per 100) for Farm B	$\frac{25.6}{570}$	= 4.5

Source of age-specific death rates: Oxender WD et al. *J Am Vet Med Assoc* 1973;162:458–460.

Comparing the age-adjusted death rates for the two farms, we see that the risk of death is greater for Farm A than it is for Farm B. This finding is consistent with the conclusion derived by comparing age-specific death rates for the two farms. This means that antibiotics did not confer a protective effect for calves on Farm A (the crude rate), but rather were used *in response to* the higher death rate and other disease problems on the farm (the adjusted rate).

5.5.2 Rate Adjustment for Other Factors

A variety of other confounding factors may bias the comparison of groups. Two of the most common in veterinary medicine are breed and sex. Furthermore, age/breed- and age/sex-specific and adjusted rates can be computed and compared as was done previously for age alone. Cause-specific morbidity and mortality rates may be stated for the entire population or for any age, breed, or sex subgroup.

Although adjusted rates can be very useful in making comparisons, the first step in examining comparative morbidity/mortality should always be to carefully examine the subgroup-specific rates for any interesting differences or changes (Gordis, 2014). Consider, for example, the data in Table 5.6 comparing the prevalence of asthma among veterinary practitioners over the 6-year period from 2006 to 2012, expressed in terms of male and female subgroups (Schelkle et al., 2017). In 2006, the prevalence of asthma in men was 58% of that for women, whereas in 2012, the prevalence had reversed so that men were 67% more likely to be afflicted with asthma. Despite this change

TABLE 5.6
Subgroup-Specific and Crude Prevalence Rates of Asthma Reported by Veterinary Practitioners in 2006 and 2012

	2006		2012	
	PAR	Asthma	PAR	Asthma
Men	304	3.9%	286	7.0%
Women	208	6.7%	308	4.2%
All	512	5.1%	594	5.6%

Source: Schelkle M et al. *Int Arch Occup Environ Health* 2017; 90: 639–643. With permission.

Abbreviation: PAR, population at risk

in subgroup-specific rates, the crude prevalence rates for 2006 (5.1%) and 2012 (5.6%) were not statistically different. This was due in part to the reversals of the gender distribution of the population at risk (PAR) for each subgroup over the study period. Thus, the crude rates hide potentially important differences in subgroup distribution and morbidity. The authors also evaluated gender differences in age, years in the profession, smoking status, and practice type, any of which alone or in combination might influence measured disease outcomes. When a number of variables can influence an outcome, multivariate analysis is appropriate. Multivariate analysis is discussed in detail in Chapter 6, "Risk Assessment and Prevention."

5.5.3 The Choice of a Standard Population

The choice of a standard population is relatively unimportant if the specific rates in one group are consistently lower than or equal to those in the other group. On the other hand, if disease rates, for example, favor younger animals in one group and older animals in another, then either group can be made to appear to have lower age-adjusted mortality rates, depending on the age distribution of the standard population. If a standard population is chosen so that it contains a large proportion of young animals, the group having the lower rates in young animals will have the lower standardized mortality. If a standard population contains a large proportion of older animals, the group having lower age-specific rates among older animals will have a low age-adjusted mortality rate. In these instances, rate adjustment or standardization may not provide more information beyond that obtained by simple comparison of specific rates (Schwabe et al., 1977).

5.5.4 When to Adjust Rates

Rates are adjusted in order to remove the effect of a factor that may confound a comparison. However, it is important to remember that the overall crude rates represent the actual events. Adjusted rates are not "real" rates in the populations being compared, because the numerical values of adjusted rates depend on the choice of the standard population used in carrying out the adjustment. An adjusted rate gives an accurate comparison, but does not reveal the underlying raw data, which are shown by the crude rate (Gordis, 2014).

Although the presence of a confounding factor is the primary condition for rate adjustment, three additional conditions must be met to justify adjusting rates;

1. A comparison is to be made (not a single population).
2. The event or characteristic of interest is defined for purpose of analysis as a rate or proportion.
3. The comparison involves overall rates (not specific rates).

5.5.5 The Uses of Incidence and Prevalence

Incidence provides a measure of the likelihood of something happening. This could be the likelihood of contracting or recovering from disease, or the duration of a disease-free state following treatment. Incidence is, therefore, the preferred statistic for expressing risk or predicting the future course of disease.

Prevalence is a measure of the status of a population at a given point in time. Because of its relationship to the predictive value of diagnostic tests, prevalence should be considered when choosing a test and interpreting its results. It is also useful in evaluating the importance of a risk factor at the population level. A factor that is associated with a high risk of disease may not be important if it is present in only a fraction of the population.

Incidence and prevalence are especially useful when used to make comparisons. Incidence and prevalence measurements are fundamental to identifying the cause during outbreak investigations.

REFERENCES

AAVMC. *Annual Data Report 2017–2018*. Washington, DC: Association of American Veterinary Medical Colleges; 2018. Full text available at: http:/www.aavmc.org/About-AAVMC/Public-Data.aspx

Armstead WW. Analytical epidemiology and veterinary economics. In Schwabe CW (ed). *Veterinary Medicine and Human Health*, 3rd ed. Baltimore: Williams & Wilkins; 1984.

AVMA. Student enrollment (1985–1986) favors women. *J Am Vet Med Assoc* 1986;188:573–575.

Buyuk F, Celebi O, Akca D et al. Estimated apparent and true prevalences of paratuberculosis in sheep herds of the Kars region in Northeastern Turkey. *Veterinarni Medicina* 2014;59:331–335. Full text available at: http://vri.cz/docs/vetmed/59-7-331.pdf

Cárdenas L, Melo O, and Casal J. Evolution of bovine brucellosis in Colombia over a 7-year period (2006–2012). *Trop Anim Health Prod* 2018;50:19–27.

Davies JA, Fransson BA, Davis AM et al. Incidence of and risk factors for postoperative regurgitation and vomiting in dogs: 244 cases (2000–2012). *J Am Vet Med Assoc* 2015;246:327–335.

Fletcher RH, Fletcher SW, and Fletcher GS. *Clinical Epidemiology—The Essentials*, 5th ed. Baltimore: Lippincott Williams & Wilkins, Baltimore; 2014.

Gordis L. *Epidemiology, 5th ed*. Philadelphia: Elsevier-Saunders; 2014.

Oxender WD, Newman LE, and Morrow DA. Factors influencing dairy calf mortality in Michigan. *J Am Vet Med Assoc* 1973;162:458–460.

Rogan WJ and Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 1978;107:71–76.

Schelkle M, Braun J, Jörres R et al. Respiratory allergies among veterinarians: Two cross-sectional surveys from 2006 to 2012. *Int Arch Occup Environ Health* 2017;90:639–643.

Schneider LG, Stromberg ZR, Lewis GL et al. Cross-sectional study to estimate the prevalence of enterohaemorrhagic *Escherichia coli* on hides of market beef cows at harvest. *Zoonoses Public Health* 2018;65:625–636.

Schwabe CW, Riemann HP, and Franti CE. *Epidemiology in Veterinary Practice*. Philadelphia: Lea and Febiger; 1977.

Tucker TR, Aly SS, Maas J et al. Investigation of *Anaplasma marginale* seroprevalence in a traditionally managed large California beef herd. *Vet Med Int* 2016;6186078. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5021891/

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 5.1: For the 2017–2018 academic year, the proportion of female veterinary students in U.S. colleges of veterinary medicine was approximately 80%. It follows that the proportion of male students was approximately 20%. The ratio of female to male students would be 80%–20% = 4 to 1, or 4 times as many female students as male students.

Answer 5.2: In this example, we have no information on when *B. abortus* infections were acquired or the duration of infection. The rate thus represents the likelihood of being a case, rather than becoming a case. Although *B. abortus* diagnoses are reported over a period of time, this is still a measure of disease prevalence because we do not know what proportion of uninfected cattle acquired the infection each year, only when the condition was first diagnosed. Infections detected in one year could have been contracted the previous year, after sampling and testing were performed. Although we cannot calculate the actual incidence, the trend in disease prevalence does suggest that the incidence of bovine brucellosis has not changed over the 7-year study period.

Answer 5.3: The authors suggest that a prospective randomized controlled clinical trial evaluating the efficacy of various preventive measures be conducted. This would facilitate a focus on cause-effect relationships without the interference of confounding variables.

Answer 5.4: Several factors, other than the geographic origin of cattle, could have influenced the difference in period prevalence at the two packing plants. The authors point out that there were many unmeasured differences between the two plants related to management, physical facilities, and employee practices that could have influenced the likelihood of EHEC contamination of cattle hides.

Answer 5.5: The calculation is as follows:

$$(8.17 \text{ cases} \div 1000 \text{ cow days}) \times (90 \text{ days/year}) \times (100 \text{ cows}) \times (1 \text{ year}) = 73.53 \text{ cases/100 cows}$$

The estimated prevalence would be 73.53%, well above the 60% benchmark proposed by the authors to prevent an anaplasmosis outbreak.

Answer 5.6: The discrepancy between apparent and true prevalence estimates in Table 5.3 are the result of correcting for the low test sensitivity of the ELISA test kit used by the authors. With such a low test sensitivity, there would be many false negatives. Substituting progressively higher values for test sensitivity into the equation reduces the difference between the two estimates. Increasing test specificity has little effect.

6 Risk Assessment and Prevention

6.1 RISK FACTORS AND THEIR IDENTIFICATION

An understanding of the concept of risk is fundamental to an understanding of the subsequent chapters on prognosis, treatment, and cause. The reason is twofold. First, all analyses rely on similar approaches to organizing and interpreting the data. Second, the statistical approach to proving that relationships exist is similar. The previous chapter focused on rates and proportions. In this chapter we will also use ratios to study associations between risk factors and outcomes.

Factors that are associated with an increased likelihood of an event occurring (such as disease) are called **risk factors**. Exposure can take place at a point in time, as when an individual comes in contact with an infectious agent or receives a drug, or may also be ongoing, like the risk of mosquito exposure for heartworm infection or cryptorchidism for testicular neoplasia.

Risk factors for many animal diseases are poorly defined or unknown and only come to light through the systematic study of naturally or spontaneously occurring cases. Clinical studies in which the researcher gathers data by simply observing events as they happen, without playing an active part in what takes place, are called **observational studies**. They are contrasted with **experimental studies**, in which the researcher determines which individuals are exposed or not exposed to the factor being investigated. Although experimental studies are more scientifically rigorous, observational studies are the only feasible way of studying most questions of risk.

Observational studies are subject to many more potential biases than are experiments. Observational study designs must minimize unwanted differences between exposure groups in order to mimic, as closely as possible, an experiment.

Observational studies are subject to many more potential biases than are experiments.

6.2 FACTORS THAT INTERFERE WITH THE ASSESSMENT OF RISK

Many risks are obvious enough that their impact on animal health can easily be documented. Exposure to pathogenic organisms and their vectors, acute toxins, or environmental stresses associated with weather extremes or transportation are recognized as major risk factors for disease. For many diseases, however, the risks are not as readily discernible, and individual clinicians are seldom in a position to assess their possible importance. Some of the reasons for this follow.

1. *Long latency*: For many conditions, the time interval between exposure and development of an outcome is too long to be perceived by a practitioner. Examples are environmental hazards such as pollutants or nutritional deficiencies, and sequelae of certain infectious diseases that may not appear until long after recovery from the initial disease, such as Lyme arthritis.
2. *High prevalence of risk factors or disease*: If a disease is relatively common among all members of a population, and some of the risk factors for it are already known, it becomes difficult to distinguish a new risk factor from the others. The effects of chronic or widespread

risk factors on animal health and production may be easily misinterpreted as the norm until they are compared with unexposed animals.

3. *Low incidence of disease*: Diseases of low incidence do not provide enough cases to prompt a practitioner to suspect that a cause-effect relationship may exist. For example, Bellumori et al. (2013) reported that approximately 30% of 90,004 mixed breed and purebred dogs examined at the University of California, Davis veterinary medical teaching hospital from 1995 to 2012 were afflicted with at least 1 of the 24 most common genetic disorders. However, the risk of occurrence of a genetic disease in any particular individual in the general canine population is usually very low. The genetic heterogeneity of outbred animals and possible polygenic nature of inherited disorders contribute to a relatively low incidence of any particular genetic defect in the population as a whole. Research into genetic diseases and possible breed predispositions is data-intensive and requires large numbers of individuals to prove an association.
4. *Small risk from exposure*: As the amount of risk conferred by a factor decreases, a larger number of subjects will be required to confirm the relationship.
5. *Multiple causes*: Many diseases exist as complexes. Examples are shipping fever, neonatal mortality, and the postpartum dysgalactia syndrome (PPDS) in sows. For these diseases, no single cause can be identified. Rather, a combination of factors acting synergistically appears to be responsible for the disease syndrome.

EXAMPLE 6.1: WHAT PROPORTION OF THE CANINE POPULATION IS AT RISK OF DEVELOPING A HERITABLE DISEASE?

Background: Nearly 700 inherited disorders and traits have been described in the domestic dog, one of the most genetically well-studied species after humans (Donner et al., 2018). Knowledge on the genetic epidemiology of disorders in the dog population has implications for both veterinary medicine and sustainable breeding. Limited data on frequencies of genetic disease variants across breeds exist, and the disease heritage of mixed breed dogs remains poorly explored to date. Advances in genetic screening technologies now enable comprehensive investigations of the canine disease heritage and generate health-related big data that can be turned into action.

Objectives: In an effort to describe the frequency and distribution of genetic markers for canine heritable diseases, Donner et al. (2018) conducted population screening of genetic variants implicated in Mendelian disorders of dogs to identify the most prevalent, and rare, disease susceptibility variants across the general dog population while providing the first extensive snapshot of the mixed breed disease heritage.

Study Design: Cross-sectional.

Methods: Genetic analyses were carried out on DNA extracted from owner-collected, non-invasive cheek swab samples, or from blood/cheek swab samples collected at certified veterinary clinics in accordance with international standards for animal care and research. The DNA of 83,220 mixed breed dogs and 18,102 purebred dogs of various breeds and geographical background was screened for 152 known Mendelian disease variants, covering the majority of the current 212 canine entries in the Online Mendelian Inheritance in Animals (OMIA) database. A large-scale DNA screening technique was used to identify genetic variants implicated in Mendelian disorders of canines.

Results: The investigators found that 40.5% (2 in 5) of all dogs carried at least one of the tested disease variants in their genome in either the hetero- or homozygous state. Most disease variants are shared by both mixed breeds and purebreds, while breed- or line-specificity of

others is strongly suggested. Mixed breed dogs were more likely to carry a common recessive disease variant, but typically in a harmless heterozygous state. In contrast, purebreds were more likely to be "genetically affected" with one of the disorders due to homozygosity for a disease variant. The authors interpret this finding as providing DNA-based evidence for hybrid vigor in mixed breed dogs. The genetic presence of 22 disease variants was discovered in at least one additional breed in which they were previously undescribed.

Conclusions and Significance: The authors conclude that risk variants for genetic disorders are prevalent in the general dog population and confirm that mixed breed dogs may suffer from many of the same medical conditions as purebreds. Mixed breed dogs were more likely to carry a common recessive Mendelian disorder, while purebred dogs were more likely to be genetically at risk of one. Some mutations likely manifest similarly independently of breed background; however, the authors emphasize the need for follow up investigations in each case. This study provides unique insight into the genetic epidemiology of canine disease risk variants, and their relevance for veterinary medicine, breeding programs, and animal welfare. In support of this effort, the authors have created *MyBreedData* (https://www.mybreeddata.com/), an online updated canine inherited disorder prevalence database based on the generated data as a free tool for breed and kennel clubs, breeders, and the veterinary and scientific community.

FOLLOW-UP QUESTION 6.1

How can veterinary practitioners and researchers contribute to and use the findings of this study and the *MyBreedData* resource? See Answer 6.1 at the end of this chapter.

6.3 USES OF RISK

1. *Prediction*: Risk is useful for estimating the likely future incidence of disease among comparable individuals. While risk for groups of individuals can be predicted rather well in this way, it is not possible to be precise about risk to any one individual in the group.
2. *Diagnosis*: The presence of a risk factor in an individual increases the likelihood that an associated disease is present and the positive predictive value of diagnostic tests for that disease. If the association between a risk factor and disease is strong, the absence of the risk factor can be used to rule out the disease. Thus, knowledge of risk factors and their associated diseases is useful for screening patients and generating a differential list.
3. *Cause*: Risk factors are frequently identified because they exhibit a statistically significant association with a disease. In some cases, this association is causal. In others, the risk factor is merely an "innocent bystander," confounded with a causal factor. Because of confounding, an association may not necessarily be a cause.
4. *Prevention*: If a risk factor is also a cause of disease, its removal can be used to prevent disease, even if the disease mechanism is unknown. For example, before bacteria were identified, a 19th-century physician by the name of John Snow found an increased rate of cholera among people drinking water supplied by a particular company in London, England. He stopped a local cholera epidemic by cutting off that supply of contaminated water (Schwabe et al., 1977). He was unaware, however, of the specific cause of the disease. The concept of cause and its relationship to prevention is discussed further in Chapters 10 and 12.

If the association between a risk factor and disease is strong, the absence of the risk factor can be used to rule out the disease.

6.4 COMPARISON OF RISKS

Several study designs and analytical techniques can be used to explore the association between presumed risk factors and outcomes. The choice of analytical technique depends, in part, on the study design employed. Regardless of the approach, results are usually expressed in terms of (1) the strength of the association between the risk factor and outcome and (2) the statistical significance of this association. In the following sections, we will discuss how these parameters are estimated.

6.4.1 Univariate Analysis

Univariate (or univariable) analysis is the simplest approach to exploring the association between a potential risk factor (**variable**) and an outcome of interest. A **two-by-two table**, illustrated in Figure 6.1, is often used to describe and analyze this relationship. The study design can be longitudinal or cross-sectional, but the layout of the table is the same. What differs is the way in which the magnitude of risk is calculated. Longitudinal studies permit disease incidence in exposed and unexposed groups to be calculated, and magnitude of risk is expressed as **relative risk**. On the other hand, cross-sectional studies provide no information on the incidence of disease but do allow us to compare how common a risk factor is among diseased (cases) and non-diseased individuals. The resulting statistic, the **odds ratio**, provides the same information as relative risk.

The derivation of these two estimates of risk is summarized in Figure 6.1 and discussed in greater detail in the following sections on study design.

6.4.2 Multivariate Analysis

Cross-sectional studies are especially useful for testing the possible causal association between a number of potential risk factors, or variables, and an outcome of interest. The analysis can be performed by constructing a two-by-two table for each of the variables, one at a time. However, it is entirely possible that two or more variables that appear to be associated with the outcome in a simple two-by-two analysis are also related to each other, such as herd size and type of housing system, or that the magnitude of risk varies with subgroups within the population being evaluated. This is known as **confounding** and makes it difficult to determine the actual contribution of suspected risk factors to disease. In these cases, two basic approaches are available to the investigator to disentangle relationships between individual variables and outcome.

6.4.2.1 Mantel-Haenszel Stratified Analysis

When the number of variables is small, a series of two-by-two analyses can be performed with each subgroup, or stratum, within the population. By stratifying on a variable, we eliminate the effect

	Cases	Noncases	
Exposed	A	B	A + B
Not exposed	C	D	C + D
	A + C	B + D	

$$\text{Relative risk} = \frac{A/(A+B)}{C/(C+D)} \qquad \text{Odds ratio} = \frac{\dfrac{A/(A+C)}{C/(A+C)}}{\dfrac{B/(B+D)}{D/(B+D)}} = \frac{A/C}{B/D} = \frac{AD}{BC}$$

FIGURE 6.1 Two-by-two table comparing how the strength of the association between exposure and outcome is estimated from cohort versus case-control studies.

of confounding by that variable. The contribution of each subgroup to the whole is weighted on the basis of its relative abundance within the population, much like direct rate adjustment (Chapter 5). This is known as a Mantel-Haenszel stratified analysis and yields an adjusted measure of risk. More detailed information on Mantel-Haenszel stratified analysis and its use to test for confounding and interaction can be found in Chapter 12.

More recently, multivariate logistic regression has replaced Mantel-Haenszel analysis as the method of choice to control for confounding while simultaneously evaluating the role of multiple suspected risk factors upon outcome.

6.4.2.2 Multivariate Logistic Regression Analysis

Multivariate, or multivariable, logistic regression (MLR) is used to assess the contribution of each of a number of potential risk factors (*X* variables) to a dichotomous outcome, such as disease or death (the *Y* variable), while controlling for confounding. The goal of a multiple logistic regression is to find an equation that best predicts the probability of the *Y* variable as a function of the *X* variables. Multiple logistic regression is especially useful if there are multiple potential predictor variables (risk factors) for an outcome, such as the contribution of a number of management factors to calf mortality. Rather than performing a series of stratified two-by-two tables (Mantel-Haenszel stratified analysis) for each subgroup, MLR is performed by constructing and solving a logistic regression equation in which the relative contribution of each risk factor is represented as an exponent. The resulting equation provides information on the magnitude and statistical significance of each variable's contribution to the outcome of interest. The statistical significance of adding or removing each *X* variable upon the overall prediction of the actual *Y* variable determines whether to include it in the model. Often a univariate analysis of each potential risk factor is performed first to identify those to be included in MLR analysis. Confounding variables that are found to be suitable for inclusion in the model from univariate analysis may "drop out" of the final MLR model because they are not found to be statistically significant. Thus, MLR is often used to reduce the number of risk factors to those that are most likely to be associated with the outcome of interest. MLR has become the method of choice for distinguishing significant risk factors for disease from myriad confounding variables. See McDonald (2014) for a more expansive treatment of MLR and other statistical techniques utilized in clinical research.

The different approaches to group selection in studies of risk and their strengths and weaknesses are discussed in the following sections.

6.5 COHORT STUDIES OF RISK

6.5.1 True Cohort Study Designs

Cohort studies, also known as longitudinal or prospective studies, involve the assembly of a group of individuals (the **cohort**) that have something in common and following them over time to detect occurrences of the outcome of interest. Cohort studies follow the same logic as the clinical question, e.g., if the subjects are exposed to a risk factor, do they contract the disease? Cohort studies are the only way of establishing incidence directly, and they can reveal the relationship between exposure and many possible outcomes.

The duration of a cohort study should be consistent with the natural history of the disease being studied. If the study is terminated too early, many cases may not yet have become detectable or run their course. Ideally, all members of the cohort study should be followed for the entire follow-up period. The study group may be assembled in the present (concurrent cohort) or from past records (historical cohort) based on any of a number of criteria. Some examples of how cohorts are used in clinical research are listed in Table 6.1. Examples of concurrent and historical cohort studies follow. An example of the use of both historical and concurrent cohorts in a clinical trial is described in Example 8.2 (Chapter 8).

TABLE 6.1
Cohorts and Their Uses

Characteristic in Common	To Assess Effect of	Example
Time	Duration of exposure	Effect of duration of dystocia in cows upon calf survival (Example 6.3)
Date of birth	Calendar time	Risk of dogs contracting leptospirosis based on season (Example 11.1)
Exposure	Etiologic agent or risk factor	Effect of low serum IgG level upon newborn calf survival (Example 6.2)
Disease	Prognosis	Prognosis for cats presenting with sepsis (Example 6.5)
Treatment	Therapeutic intervention	Effectiveness of ureteral stenting for treatment of feline ureteral obstruction (Example 8.2)

6.5.1.1 Concurrent Cohort Studies

In a **concurrent cohort study**, the study group is assembled in the present and followed into their future. This study design usually requires periodic examination of members of the cohort to record new occurrences of the event of interest.

EXAMPLE 6.2: HOW IMPORTANT IS COLOSTRUM INTAKE FOR NEWBORN CALVES?

Background: Calves are born nearly devoid of maternal antibodies and depend on intestinal absorption through colostrum intake shortly after birth to acquire sufficient immunity to common pathogens. Colostrum management is one of the most crucial factors that affect calf health and survival; inadequate colostrum intake could result in failed transfer of passive immunity (FTPI) of immunoglobulin (Ig)G. Although failed transfer of passive immunity is commonly recognized as a main factor affecting calf health and survival, conflicting results are reported in the literature as to the strength of association between passive immunity (PI) and calf health, especially regarding enteric diseases.

Objectives: Lora et al. (2018) performed a cohort study evaluating the association between PI and health status of calves within 30 days of age under field conditions.

Study Design: Cohort study.

Methods: Blood samples were collected from 78 calves between 1 and 5 days of age on three commercial Italian dairy farms during the winters of 2014–2016. Immunoglobulin concentration, and disease and mortality occurrence were recorded during the first month of life. Additionally, fecal samples were collected from calves with scours before treatment. Blood serum samples were tested by an electrophoretic method for the assessment of immunoglobulin (Ig) concentration, whereas fecal samples were evaluated with an ELISA test for positivity to *Escherichia coli* K99, rotavirus, coronavirus, and *Cryptosporidium spp.* Calves were divided into two categories based on serum Ig concentration: having FTPI (serum Ig concentration <10.0 g/L) or not having FTPI (serum Ig concentration ≥ 10 g/L). The effect of FTPI as a predictive factor for disease and mortality occurrence, antibiotic treatments for sick calves, and specific enteropathogen infections was investigated through logistic regression (see below), including the farm in the model. Odds ratios and 95% CI were calculated for the dependent variables that were significantly affected by FTPI. A Kaplan–Meier survival plot (see Chapter 7) was used to evaluate the age (days) at first diarrhea occurrence in the 78 calves stratified by presence or absence of FTPI.

Results: Only enteric diseases occurred in calves of this study. The overall percentage of calves with FTPI was 34.6%. The investigators found that FTPI in calves less than 1 month of age was associated with a statistically significant increase in the incidence of diarrhea and mortality. The mortality rate for FTPI calves was 14.81% versus only 1.96% for calves with normal immunoglobulin levels, signaling the possible need for antibiotic therapy during this critical period. Logistic regression confirmed that calves that suffered from diarrhea or died within the first month of life had lower serum Ig concentrations than those that remained healthy or survived ($p < 0.05$). Calves with adequate PI transfer also had a 6-day delay in the age at first disease onset compared to those with FTPI ($p < 0.01$). Even when estimated on a small number of calves, those with FTPI had higher risks of enteric infections by rotavirus (odds ratio = 12; 95% CI = 1–137) and *Cryptosporidium spp.* (odds ratio = 9; 95% CI = 1–72) ($p < 0.05$).

Conclusions and Significance: In this study, the PI level influenced the occurrence of enteric diseases and mortality in calves under 1 month of age, confirming the importance of a proper colostrum provision to calf health and, consequently, to the reduction of antimicrobial use in dairy farming. However, further investigations are needed, particularly focusing on the relationship between PI and specific enteropathogen infections in calves.

FOLLOW-UP QUESTION 6.2

Based on the above findings, how many times more likely are calves with FTPI to die than calves with normal immunoglobulin levels? See Answer 6.2 at the end of this chapter.

6.5.1.2 Historical Cohort Studies

In a **historical cohort study**, the study group is assembled from past records and followed into their future, usually up to the present. The term **retrospective cohort** is also used to describe a historical cohort. The term *cohort* is used because every individual has an equal chance of being included in the study, e.g., sampling based on exposure. The term *retrospective* is used because evidence of exposure is based on past records or recall.

EXAMPLE 6.3: HOW SOON AFTER CLINICAL SIGNS OF DYSTOCIA IN CATTLE BECOME EVIDENT SHOULD VETERINARY ASSISTANCE BE REQUESTED?

Background: Cesarean section is indicated for dystocia in cows when assisted vaginal delivery is ineffective and fetotomy is not an option. It is essential that farm personnel be able to determine whether, when, and how to intervene and when to seek veterinary assistance for dams with signs of labor.

Objectives: Hiew et al. (2018) reported on the signalment, clinical signs, reproductive history, surgical management, and postoperative outcomes of beef cattle undergoing cesarean section for correction of dystocia. A secondary objective was to identify factors associated with an unfavorable outcome for the fetus or dam.

Study Design: Historical cohort study.

Methods: Medical records of 303 female cattle admitted to the Veterinary Teaching Hospital of Purdue University for dystocia over a 10-year period were reviewed. A subset of 173 beef cattle with dystocia that underwent cesarean section (the historical cohort) was selected for analysis. Survival information was available for 165 (92.7%) of calves, of which 62 (37.6%) were born dead or died ≤24 hours after birth.

In order to isolate the relationship between duration of dystocia and calf survival from all of the other dam- and calf-related confounding variables, a multivariable logistic regression model was used to predict calf survival 24 hours after cesarean section as a function of the duration of dystocia. Other variables ("predictors") included in the model were administration of epidural analgesia to the dam, calf sex, and dam age. For the prediction of survival, adjusted odds ratios (ORs) and 95% confidence intervals were estimated simultaneously for predictors added in a forward stepwise fashion to the logistic regression equation on the basis of a value of $p < 0.05$ for the maximum likelihood estimate.

Results: Results confirmed that duration of dystocia provided the best predictor of calf survival. A significant ($p < 0.001$) association was identified between duration of dystocia and calf survival, with the percentage of surviving calves decreasing as the duration increased. At the optimal test cutoff point (duration of dystocia <6 hours), the sensitivity and specificity of the model for predicting calf survival were 53% and 96%, respectively. The positive likelihood ratio (sensitivity/1-specificity) was 13.3.

Conclusions and Significance: Based on the high overall mortality rate for calves (37.6%) and the results of MLR analysis, the authors encourage beef producers to seek veterinary assistance whenever clinical signs of dystocia are evident, preferably within 6 hours after the onset of parturition.

FOLLOW-UP QUESTION 6.3

The sensitivity and specificity of the MLR-derived model for predicting calf survival, based on duration of dystocia <6 hours, were 53% and 96%, respectively. The positive likelihood ratio was 13.3. How can these measures of test performance be used for ruling in or ruling out the likelihood of calf survival? See Answer 6.3 at the end of this chapter.

Note that in both of the preceding examples, calf morbidity and mortality shortly following birth were evaluated. However, in the first example, the cohort was formed in the present, at the time calves were born, whereas in the second example, the cohort was assembled based on births that occurred years earlier. Both studies looked "forward" into the calves' futures from the point of exposure to a risk factor (birth), and therefore were able to report incidence of positive and negative outcomes.

Cohort studies, also known as prospective studies, involve the assembly of a group of individuals that have something in common and following them over time to detect occurrences of the outcome of interest.

6.5.2 Comparing Risks in Cohort Studies

Incidence is a basic expression of risk. It is the number of new events (usually disease) arising in a defined population over a given period of time. Incidence is especially useful for evaluating the relationship between presumed risk factors and disease. Several measures, called **measures of effect**, can be estimated from incidence data.

6.5.2.1 Relative Risk

Relative risk (RR), or **risk ratio**, is calculated by dividing incidence in individuals exposed to a risk factor by incidence in nonexposed individuals. Relative risk can range from zero to infinity. If no additional risk is associated with exposure, then both incidences should be equal and the ratio would be equal to one.

Relative risk is an index of the strength of the association between a risk factor and disease, but tells us nothing about the absolute magnitude of that risk. For this we must calculate the attributable risk.

6.5.2.2 Attributable Risk

Attributable risk (AR), also known as **risk difference**, is calculated by subtracting incidence among those not exposed to a risk factor from incidence among exposed individuals. Since subtraction removes background incidence, attributable risk is the additional incidence of disease attributable to the risk factor itself. Considered another way, it is the disease incidence that would not occur had the risk factor not been present.

The difference between relative risk and attributable risk can be appreciated if we consider that a tenfold reduction in incidence among both exposed and unexposed would result in a tenfold reduction in attributable risk, but would have no effect upon relative risk.

6.5.2.3 Population Attributable Risk

Relative and attributable risks provide information on the contribution of risk factors to the overall rates of disease in exposed individuals. However, neither tells us how much a risk factor contributes to the overall rate of disease in the population or herd. This information would be useful in deciding which risk factors are important and which are trivial in the overall incidence of a particular disease in a population, or which risks are associated with the greatest economic loss.

Population attributable risk (PAR) is estimated by multiplying the attributable risk for a particular risk factor by the prevalence of that risk factor in the population. It provides a measure of how much a risk factor contributes to disease incidence at the population level. A relatively weak risk factor that is quite prevalent could contribute more to overall disease incidence than a stronger risk factor that is rarely present.

6.5.2.4 Population Attributable Fraction

We may also wish to know what fraction of disease occurrence in a population is associated with a particular risk factor. This is called the **population attributable fraction (PAF)** and is estimated by dividing the population attributable risk by the total incidence of that disease in the population. The population attributable fraction permits us to predict the proportion of cases of a particular disease that will be eliminated through control of a particular risk factor. If all cases are associated with the risk factor being measured, then the population attributable fraction would be 1.00, or 100%.

Table 6.2 compares the above measures of effect for the risk of disease. The significance of these expressions of risk can be appreciated in the following example.

TABLE 6.2
Measures of Effect in Studies of Risk of Disease

Expression	Clinical Question	Calculation[a]
Relative risk (risk ratio)	How many times more likely are exposed individuals to become diseased relative to unexposed?	RR = IE ÷ Ie
Attributable risk (risk difference)	What is the incidence of disease attributable to exposure?	AR = IE – Ie
Population attributable risk	What is the incidence of disease in a population associated with the occurrence of a risk factor?	ARp = AR × P
Population attributable fraction	What fraction of disease in a population is attributable to exposure to a risk factor?	AFp = ARp ÷ RT

Source: Fletcher RH et al. *Clinical Epidemiology—The Essentials*, 1st ed. Baltimore: Williams & Wilkins, Baltimore; 1982. With permission.

[a] Where IE = incidence in exposed individuals; Ie = incidence in nonexposed individuals; P = prevalence of exposure to a risk factor; and RT = total incidence of disease in a population.

EXAMPLE 6.4: WHAT IS THE FARM-LEVEL IMPACT OF FAILED TRANSFER OF PASSIVE IMMUNITY TO NEWBORN CALVES?

The following analyses incorporate data from the study by Lora et al. (2018) on FTPI in calves described in Example 6.2. Calf exposure (FTPI status) and mortality data from that study has been inserted into Table 6.3 for the calculation of **simple risks**. In Table 6.4, these values are used to calculate the **compared risks** associated with FTPI in newborn calves.

TABLE 6.3
Two-by-Two Table Analysis of Data from Lora et al. (2018)

	Deaths	Survivors	Total
Exposed (to low immunoglobulin levels)	(a)	(b)	27
	4	23	
Not exposed (to low immunoglobulin levels)	(c)	(d)	51
	1	50	
Total	5	73	78

Incidence in exposed = (4 ÷ 27) = 14.81%
Incidence in unexposed = (1 ÷ 51) = 1.96%
Prevalence of low Ig = (27 ÷ 78) = 34.6%
Overall mortality rate = (5 ÷ 78) = 6.41%

TABLE 6.4
Interpretation of Measures of Effect Analysis: Suboptimal Immunoglobulin Levels in Calves; Data from Lora et al. (2018)

Simple risks

Incidence of calf losses among low gamma globulin group = 14.81%
Incidence of calf losses among remaining calves = 1.96%
Prevalence of low gamma globulin levels in all calves = 34.6%
Incidence of calf losses = 6.41%

Compared risks

Relative risk (14.81 ÷ 1.96) = 7.56
Attributable risk (14.81 – 1.96) = 12.85%
Population attributable risk (12.85 × .346) = 4.45%
Population attributable fraction (4.45 ÷ 6.41) = 69.4%

FOLLOW-UP QUESTION 6.4

How would you communicate the results of the **compared risks** analysis in Table 6.4 to a producer? See Answer 6.4 at the end of this chapter.

6.5.3 Limitations of Cohort Studies

Since concurrent cohort studies are conducted in the present, they permit the collection of any data required for the specific purposes of the study. In contrast, data for historical cohort studies are often limited to what was recorded in medical or herd records. Vital information may be difficult or impossible to obtain. Historical cohorts are useful when it would take so long for an event to occur that the experiment would be jeopardized. For example, a study examining the risks associated with

being a veterinarian could conceivably extend beyond the lifetime of the investigators if it were conducted as a concurrent cohort study.

Regardless of the way in which a cohort study is conducted, if all individuals are identical at the time they enter into a study, and the only variable is the time over which they will be followed, then a true cohort study exists. If there is reason to believe that differences exist among individuals that may influence the outcome of the study, then a biased view of risk may result.

One of the major difficulties in cohort studies is assembly of all members of the cohort at the same time. For example, if one wished to assess the risk of testicular neoplasia in cryptorchid dogs, individuals exposed to the risk factor (cryptorchidism) may not all be available at the same point in time. This affects their follow-up period, and outcome must be expressed as incidence density. Even if all individuals can be assembled at the same point in time, additional difficulties may affect the validity of cohort studies.

Cohort studies also lack the controls inherent in laboratory experiments. Additional factors such as diet, housing, management, and exposure to other animals are difficult to control and may influence the outcome of cohort studies. Diseases of low incidence present a special problem. The number of animals that must be assembled and followed to assure that a sufficient number of cases will arise may make a cohort study impractical. An alternative approach, the **cross-sectional study**, is discussed later in this chapter.

6.5.4 Case Series

Concurrent and historical cohorts are sometimes referred to as **true cohorts**, since they are studied from the point at which subjects are first exposed to a risk factor or at the onset of disease. Sometimes this is not possible, and the cohort must include individuals at any stage of their disease. This assemblage of individuals is most commonly found in studies described as **case series**. Case series are purely descriptive studies for which the objective is to characterize or summarize the typical clinical course of a particular condition or disease through to some outcome (Lefebvre and Matushek, 2018). The data can be obtained either retrospectively, through review of medical records, or prospectively, as cases are encountered. In either case, each subject is recruited into the study after a given period of exposure or disease.

EXAMPLE 6.5: WHAT ARE THE CLINICAL AND LABORATORY FINDINGS ASSOCIATED WITH SEPSIS IN CATS? WHAT IS THE PROGNOSIS?

Background: Sepsis (systemic inflammatory response syndrome—SIRS) is a severe, potentially fatal clinical syndrome that results from a systemic inflammatory response to infection by bacteria, viruses, protozoa, fungi, or helminths. In cats, sepsis is associated with substantial illness and high mortality rates. Although naturally occurring sepsis is well studied in human patients, studies in cats with sepsis are limited, and there is no consensus regarding the diagnosis of SIRS in cats.

Objectives: Klainbart et al. (2017) conducted a study to characterize clinical and laboratory findings in cats with naturally occurring sepsis, emphasizing hemostasis-related findings, and evaluate these variables for associations with patient outcomes. The authors hypothesized that cats with sepsis would sustain inflammation-induced hemostatic derangement.

Study Design: Case series.

Methods: Thirty-one cats (18 males and 13 females) were enrolled in the study over a 3-year period when presented at the Emergency and Critical Care Department of Hebrew University Veterinary Teaching Hospital. Data collected included history; clinical signs; results of hematologic, serum biochemical, and hemostatic tests; diagnosis; and outcome. A

randomly-selected group of 33 healthy cats served as a comparison group for initial hemostatic variable analysis but were not followed prospectively. Differences between cats with and without sepsis and associations between variables of interest and outcome were explored.

Results: The sepsis group included cats with pyothorax (n = 10), septic peritonitis (7), panleukopenia virus infection (5), bite wounds (5), abscesses and diffuse cellulitis (3), and pyometra (1). Common clinical abnormalities included dehydration (21 cats), lethargy (21), anorexia (18), pale mucous membranes (15), and dullness (15). In addition, study subjects suffered from multiple hematologic, biochemical, and hemostatic abnormalities on hospital admission, including several findings suggestive of hemostatic derangement. The 30-day survival rate of cats with sepsis was 19 of 30 (63%). Univariate analysis revealed a number of clinicopathologic factors positively and negatively associated with death. However, in the final multivariate logistic regression model, none of these associations was significant. There was no difference in mortality rates among cats with sepsis that could be attributed to different underlying diseases ($p = 0.34$).

Conclusions and Significance: The authors concluded that additional research including larger numbers of cats is needed to further investigate these findings and explore associations with outcome.

FOLLOW-UP QUESTION 6.5

The authors suggest that their findings should be interpreted with caution. What aspects of this study might have led to this interpretation? See Answer 6.5 at the end of this chapter.

Regardless of the way in which a cohort study is conducted, if all individuals are identical at the time they enter into a study, and the only variable is the time over which they will be followed, then a true cohort study exists.

6.6 CASE-CONTROL STUDIES OF RISK

The prospective approach to the estimation of risk, prognosis, and treatment outcomes relies on assembly of a large number of individuals, some of whom are exposed to a factor or an intervention, and some who are not. This approach makes for good science but does not make the best use of the unique resource most readily available to the practitioner, i.e., the clinical cases. Furthermore, the frequency of many diseases of veterinary concern is relatively low. A statistically significant cohort study of risk factors may require us to follow extremely large numbers of animals over long periods of time. This could make prospective studies of risk and prognostic factors, and treatments for these diseases, impossible.

Rather than forming cohorts with the desired characteristics (risk factors) and then waiting an unpredictable period of time for something to happen, wouldn't it make more sense to start with diseased individuals and "look backward" to compare the proportion of "cases" that were exposed to the factor(s) of interest with a comparable group of "noncases"? This approach, known as a **case-control study**, is fundamental to studies of uncommon diseases and in outbreak investigations where the practitioner must rule out a number of possible risk factors. The approach also lends itself to clinical studies of risk and prognosis using medical records.

6.6.1 Advantages of Case-Control Studies

Case-control studies lend themselves to clinical research because they take advantage of a resource that practitioners have in abundance—cases. Since case-control studies start with cases, comparisons

are not constrained by diseases of low frequency or long latency. For example, Parkin et al. (2018) estimated that the minimum practice-related accident rate for equine practitioners in the United Kingdom was approximately 5400 per 100,000 per annum, or 0.054 per individual, based on a questionnaire survey of 620 equine veterinarians. If a cohort study were designed to generate at least 100 practice-related injuries, more than 1800 equine practitioners would have to be enrolled in the study and followed for a year. The expense and logistical difficulties of such a study design would be challenging. In contrast, it would be more feasible and relatively inexpensive to assemble 100 or more cases of practice-related injuries among equine practitioners, find similar groups of equine practitioners that had not experienced any injuries, and compare frequencies of hypothesized risk factors (such as age, sex, work status, and practice environment).

Another advantage of case-control studies is that large numbers of possible risk or causal factors for a disease syndrome of unknown etiology can be explored. Whereas cohort studies are designed to examine the role of a limited number of causal factors, the number of causal factors that a case-control study can consider is much greater, provided of course that data on the frequency of the suspected causal factors can be obtained from medical records or through interviewing techniques. The case-control design lends itself to "fishing expeditions."

Advantages of case-control studies are (1) cases can be identified unconstrained by the natural frequency of disease, (2) studies are unaffected by latency of disease, and (3) large numbers of possible risk or causal factors can be explored.

6.6.2 Comparing Risks in Case-Control Studies

In the cohort approach, sampling is based on exposure, whereas in the case-control approach, sampling is based on outcome. Both cohort and case-control study designs measure frequency, but in cohort studies, the frequency of different outcomes is measured, whereas in case-control studies, the frequency of the presumed causal factors is measured. As opposed to the cohort study, evidence of exposure in case-control studies usually relies on memory and the availability and completeness of medical or herd records. It is the past, not the present, that is important, and therein lies a potential for bias in case-control studies.

In the cohort approach, sampling is based on exposure, whereas in the case-control approach, sampling is based on outcome.

6.6.3 The Odds Ratio

Since the case-control study begins with the selection of cases, we have no data on the size of the population at risk and consequently the incidence of disease. Mathematically, adding cases or controls to a two-by-two table would alter the value obtained for incidence, which doesn't make biological sense. It is, therefore, not possible to calculate relative risk in the usual way. However, it is possible to obtain an **estimate** of relative risk in another way. The **odds ratio**, defined as the odds that a case is exposed divided by the odds that a control is exposed, provides a measure of risk for case-control studies that is conceptually and mathematically similar to the relative risk (Figure 6.1). The meaning of the odds ratio is analogous to the relative risk obtained in cohort studies, e.g., the higher the odds ratio, the stronger the association between exposure and disease.

The meaning of the odds ratio is analogous to the relative risk obtained in cohort studies, e.g., the higher the odds ratio, the stronger the association between exposure and disease.

6.6.4 Bias in Case-Control Studies

There are three major sources of bias in case-control studies: (1) the selection of groups, (2) measurement of exposure, and (3) presumed temporal relationships.

6.6.4.1 Bias in Selecting Groups

Case-control studies are designed to test whether there is a significant difference between cases and controls with regard to exposure to a suspected risk factor. It is essential, therefore, that the selection process ensure that both groups have an equal likelihood of being detected as cases if they develop the condition of interest. This will facilitate the detection of risk factors that are significantly associated with disease. Bias in selection of groups can be reduced by (1) matching cases with one or more controls for factors already known to be related to disease, and (2) choosing more than one control group, preferably from a different geographic location.

6.6.4.2 Bias in Measuring Exposure

Measurement bias may occur when the presence of the outcome affects the owner's recollection of the exposure (recall bias), or the measurement or recording of the exposure. These sources of bias may be reduced by (1) using alternative sources for the same information and (2) concealing the specific purpose of the study from interviewers and interviewees.

6.6.4.3 Presumed Temporal Relationships

Although case-control studies are often considered to be longitudinal, the fact remains that sampling is cross-sectional, i.e., occurs at one point in time. Unless presumed risk or causal factors are innate characteristics of the individual (as breed or sex), it may be difficult to document the temporal relationship between the risk factors being examined and the outcome of interest.

EXAMPLE 6.6: CAN HAVING A BIRD FEEDER BE HAZARDOUS TO YOUR HEALTH?

Background: Psittacosis, caused by the bacterial agent *Chlamydia psittaci*, is a zoonotic disease carried by at least 460 different species of birds, the principal reservoir. Transmission from birds to humans occurs mainly through inhalation of the bacteria aerosolized from respiratory secretions or dried feces of infected birds. People of all ages are susceptible to psittacosis, but it is more commonly reported among adults. Those who have contact with pet birds and poultry, including people who work in bird-related occupations such as aviary and pet employees, poultry workers, and veterinarians, are at increased risk. Psittacosis is a notifiable disease in Sweden, where an average of eight cases per year were reported between 2002 and 2012.

Objectives: An outbreak of 25 cases in early 2013 drew attention to the disease and prompted the following study that explored the relationship between exposure to wild and domestic birds and the risk of contracting the disease.

Study Design: Case-control study.

Methods: Chereau et al. (2018) conducted a matched case-control study of all domestically acquired psittacosis cases reported between December 2014 and April 2016 to *SmiNet*, the Swedish reporting system for notifiable diseases. Twenty-nine cases were matched to 74 controls selected from a commercial population register on the basis of age, sex, residence postal code, and season of onset of matched cases. All participants completed a questionnaire that detailed the nature and extent of exposures to wild and domestic birds. Potential risk factors ($n = 16$) were initially evaluated through univariate analysis by calculating matched

odds ratios with 95% confidence intervals (95% CI) and *p*-values. All risk factors with a *p*-value <0.2 in the univariate analysis (n = 11) were included in a multivariate conditional logistic regression model. Variables with a *p*-value >0.05 were then excluded from the multivariate model in a stepwise fashion by performing a likelihood-ratio test at each step to ensure that removal of the variable did not affect the fit of the model.

Results: The final multivariate model found that two independent risk behaviors from two different avian sources were significantly associated with an increased likelihood of contracting psittacosis: cleaning a wild bird feeder (OR = 18.95; 95% CI: 2.11 – 170.03, *p*-value = 0.009), and owning domestic birds, e.g., chickens, ducks, turkeys, pigeons (OR = 5.55, 95% CI: 1.16–26.61, *p*-value = 0.032).

Conclusions and Significance: The results suggest that exposure to bird feces, for example, when cleaning a wild bird feeder, was the main route of transmission in this outbreak. Following this study, the Public Health Agency of Sweden encouraged the use of bird feeders with a design limiting fecal accumulation, avoiding the production of dust from dried bird feces when cleaning contaminated surfaces, and washing hands afterwards.

FOLLOW-UP QUESTION 6.6

How would you interpret the results of this study for a client who asks about their risk of contracting psittacosis? See Answer 6.6 at the end of this chapter.

6.7 PREVALENCE SURVEYS OF RISK

A **prevalence survey** is a cross-sectional design that bears some similarities to both cohort and case-control approaches. As in the cohort study, the prevalence survey begins with a **defined population**. However, rather than measuring an outcome, the investigator divides the population into cases and noncases and then measures the prevalence of the putative risk factor in each group, as in the case-control approach.

6.7.1 Comparing Risks in Prevalence Surveys

In a prevalence survey, we can be certain that cases and noncases came from the same population, but the exposure history must be reconstructed from interviews or medical records. Additionally, the cases include only those detected, or prevalent, during the examination. As the sampling strategy is essentially random, the resulting relative risk or odds ratio estimates would remain relatively unchanged if additional individuals are added to the study. However, since incidence is not being measured, it is preferable to use the odds ratio to express risk in prevalence surveys. An exception is a prevalence survey conducted during the course of an outbreak investigation. In this case, incidence (**attack rate**) is actually being measured and it is possible to define the temporal sequence between exposure and disease. Examples are foodborne disease and other similar outbreaks that occur over a defined and relatively short period of time. In these cases, attack rates are calculated, so relative risk can be used.

Prevalence surveys are especially common in clinical research using medical records. Typically, the records are scanned for all cases of the condition of interest over some time interval. The prevalence of one or more suspected risk factors (age, breed, sex, etc.) among cases is then compared with prevalences for the remaining clinic population, or a defined, low-risk subpopulation, over the same period (e.g., the noncases). The strength of association of each suspected factor is expressed as an odds ratio and its statistical significance tested with the **Chi-square test**.

EXAMPLE 6.7: HOW CAN THE QUALITY OF LIFE FOR VETERINARY STUDENTS BE IMPROVED?

Background: A veterinarian's professional quality of life is influenced by a number of competing stressors and emotionally satisfying forces. During the course of their professional training, veterinary students are subjected to an immense workload and comparable forces that may impact their psychological well-being.

Objectives: In an attempt to assess the magnitude of these outcomes, and the contribution of a number of predictor variables, McArthur et al. (2017) conducted a cross-sectional online survey of Australian Veterinary Students to (1) determine the prevalence of compassion satisfaction, burnout, and secondary traumatic stress among students, and (2) investigate the association between these outcomes and several predictor variables to provide insight into possible strategies for their mitigation.

Study Design: Cross-sectional (survey).

Methods: A total of 193 (23%) usable responses out of 828 students sampled from six of the seven Australian veterinary schools were analyzed. **Predictor variables** included: (1) self-stigma—inability or reluctance to seek psychological help from a therapist, (2) coping strategies, (3) empathy—the ability to appreciate or react emotionally to another's perspective, and (4) mindfulness—enhanced attention to current experiences with a nonjudgmental attitude. **Outcome variables** included: (1) compassion satisfaction—positive feelings through helping others, (2) burnout—feeling overwhelmed and frustrated, and (3) secondary traumatic stress—exposure to and internalization of stressful experiences of others. Outcomes were assessed through participants' level of agreement with 30 statements on the Professional Quality of Life Scale, known as the ProQOL, over the preceding 30 days. Associations between the psychological predictors and outcomes were assessed through bivariate correlation and multiple regression analyses.

Results: The authors concluded that approximately 30% of veterinary students were at high risk of burnout, 24% were at high risk of secondary traumatic stress, and 21% reported low compassion satisfaction. High empathic concern, low personal distress, female gender, and employment history at a veterinary clinic were associated with high compassion satisfaction. High dysfunctional coping, low nonjudgmental and acting-with-awareness mindfulness, and lack of previous employment at a veterinary clinic were associated with high burnout. High dysfunctional coping, low acting-with-awareness mindfulness, high self-stigma, and high personal distress were associated with high secondary traumatic stress.

Conclusions and Significance: The authors proposed a number of interventions to prevent and manage burnout and secondary traumatic stress while enhancing compassion satisfaction, empathy, coping strategies, and mindfulness. The use of peer-to-peer techniques, open nonjudgmental discussion groups, and virtual patients may be useful approaches for improving veterinary student well-being.

FOLLOW-UP QUESTION 6.7

What are some of the limitations imposed by the study design upon the interpretation of the findings from this study? See Answer 6.7 at the end of this chapter.

TABLE 6.5
Comparison of Characteristics of Cohort, Case-Control, and Prevalence Survey Study Designs

Cohort	Case-Control	Prevalence Survey
Begins with a defined population at risk	Population at risk generally undefined	Begins with a defined population
Cases not selected but ascertained by continuous surveillance	Cases selected by investigator from an available pool of patients	Cases not selected but ascertained by a single examination of the population
Comparison group not exposed to risk factor but similar to at-risk group in other regards	Controls selected by investigator to resemble cases	Noncases include those free of disease at a single examination
Exposure measured before the development of disease	Exposure measured, reconstructed, or recollected after development of disease	Exposure measured, reconstructed, or recollected after development of disease
Risk or incidence of disease and relative risk measured	Risk or incidence of disease cannot be measured directly; relative risk of exposure can be estimated by odds ratio	Risk or incidence of disease cannot be measured directly; relative risk of exposure can be estimated by odds ratio

Source: Fletcher RH et al. *Clinical Epidemiology—The Essentials*, 1st ed. Baltimore: Williams & Wilkins; 1982. With permission.

6.7.2 Limitations of Prevalence Surveys

Prevalence surveys are especially useful in situations where we wish to determine which of a number of potential risk factors is associated with an outcome, as during disease outbreak investigations. Prevalence surveys are less useful for examining the role of a specific causal factor, because cases and controls are not purposely matched to control for bias. Whatever matching of cases and controls that does occur in a prevalence survey is merely a fortuitous result of their being drawn from the same population. Another problem with prevalence surveys (and cross-sectional surveys in general) is that it may not be possible to distinguish between a risk factor and a prognostic factor for a condition. In other words, a factor that does not affect disease incidence but is related to survival of the cases will be associated with disease prevalence in a cross-sectional study.

Characteristics of cohort, case-control, and prevalence survey designs are compared in Table 6.5.

6.8 BIOLOGICAL PLAUSIBILITY AND CROSS-SECTIONAL STUDY DESIGNS

A distinguishing feature of both case-control and prevalence survey designs, which contributes to their fallibility, is that subjects possess the outcome of interest at the time that the clinical findings or causal factors are measured. In some cases, temporal relationships between presumed causes and their effects are obvious, such as breed or sex predisposition to particular disease outcomes. In others the cause-effect relationship is not so clear. In these cases, the validity of the presumed temporal relationships must be based on our understanding of the mechanisms of disease, e.g., **biological plausibility**. In fact, this illustrates the mutual dependency of epidemiologic and mechanistic (or basic) research. Epidemiologic studies cannot prove with certainty that a cause-effect relationship exists, only that an association exists. Research on mechanisms of disease provides the biological basis for believing that associations are, in fact, causal. Likewise, information derived from research on mechanisms of disease cannot assume that a particular phenomenon behaves in nature as it does in the laboratory. For this, epidemiologic studies must be conducted.

REFERENCES

Bellumori TP, Famula TR, Bannasch, DL et al. Prevalence of inherited disorders among mixed-breed and purebred dogs: 27,254 cases (1995–2010). *J Am Vet Med Assoc* 2013;242:1549–1555.

Chereau F, Rehn M, Pini A et al. Wild and domestic bird faeces likely source of psittacosis transmission—A case-control study in Sweden, 2014–2016. *Zoonoses Public Health* 2018;65:790–797. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/zph.12492

Donner J, Anderson H, Davison S et al. Frequency and distribution of 152 genetic disease variants in over 100,000 mixed breed and purebred dogs. *PLoS Genet* 2018;14:e1007361. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5945203/

Fletcher RH, Fletcher SW, and Wagner EH. *Clinical Epidemiology—The Essentials*, 1st ed. Baltimore: Lippincott Williams & Wilkins; 1982.

Hiew MWH, Baird AN, and Constable PD. Clinical signs and outcomes of beef cattle undergoing cesarean section because of dystocia. *J Am Vet Med Assoc* 2018;252:864–872.

Klainbart S, Agi L, Bdolah-Abram T et al. Clinical, laboratory, and hemostatic findings in cats with naturally occurring sepsis. *J Am Vet Med Assoc* 2017;251:1025–1034.

Lefebvre SL and Matushek KJ. New guidance for reporting of case series in the JAVMA. *J Am Vet Med Assoc* 2018;252:8–9.

Lora I, Gottardo F, Contiero B et al. Association between passive immunity and health status of dairy calves under 30 days of age. *Prev Vet Med* 2018;152:12–15.

McArthur ML, Andrews JR, Brand C et al. The prevalence of compassion fatigue among veterinary students in Australia and the associated psychological factors. *J Vet Med Educ* 2017;44:9–21.

McDonald JH. *Handbook of Biological Statistics*, 3rd ed. Baltimore: Sparky House Publishing; 2014. Full text available at: http://www.biostathandbook.com/index.html

Parkin TDH, Brown J, and Macdonald EB. Occupational risks of working with horses: A questionnaire survey of equine veterinary surgeons. *Equine Vet Educ* 2018;30:200–205. Full text available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/eve.12891

Schwabe CW, Riemann HP, and Franti CE. *Epidemiology in Veterinary Practice*. Philadelphia: Lea and Febiger; 1977.

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 6.1: The identification of Mendelian genetic risk variants described in this study suggests a predisposition of carriers to particular clinical outcomes but did not document such outcomes. This relationship could be explored through any combination of follow-up owner interviews, examination of medical records and background information on individual patients, and prospective and/or retrospective studies of risk. Prospective and retrospective study designs are discussed later in this chapter.

Answer 6.2: The answer to this question is expressed in terms of **relative risk**, the incidence of disease in those exposed to a risk factor divided by the incidence in those not exposed. The authors of this study reported a mortality rate for FTPI calves of 14.81% versus only 1.96% for non- FTPI calves, yielding a relative risk of 7.56. In other words, calves with FTPI were 7.56 times more likely to die during the first 30 days of age than calves with normal immunoglobulin levels.

Answer 6.3: Based on a positive likelihood ratio of 13.3 at the optimal test cutoff point for predicting calf survival (duration of dystocia <6 hours), the authors suggest that this variable would be very good at ruling in a "diagnosis" of calf survival. Knowledge that the duration of dystocia was <6 hours would therefore be clinically helpful in supporting a decision for cesarean section in beef cattle with dystocia, despite the low test sensitivity of 0.53.

Answer 6.4: From the results in Table 6.4 we can conclude the following:

1. Calves with low serum gamma globulin levels are more than seven times as likely to die than their "normal" counterparts (relative risk).
2. Low serum gamma globulin levels are associated with more than12% additional deaths among exposed calves (attributable risk).

3. Low serum gamma globulin levels are associated with almost 4.5% deaths among all calves (i.e., the herd, population attributable risk).
4. Low serum gamma globulin levels are associated with almost 70% of all calf losses among all calves in the first month of life (population attributable fraction).

Answer 6.5: Although the authors were able to generate an extensive collection of descriptive statistics on patient findings and possible associations with outcomes, they identified several factors that limited the strength of conclusions that could be drawn from their study. These limitations included multiple etiologies of sepsis among patients, variable severity and duration of disease upon enrollment in the study, variable follow-up period, missing data and lack of serial measurements following initial sampling, limited sample size and statistical power, and inclusion of findings that were part of the case definition. These deficiencies do not reflect negatively upon the capabilities of the investigators, but rather of the challenges inherent in patient-based research.

Answer 6.6: Since groups in this case-control study were formed on the basis of outcome rather than exposure, the study provides no information on disease incidence, e.g., the risk of contracting psittacosis. However, you could state that the most likely source of infection for psittacosis is through cleaning bird feeders and working with domestic birds such as chickens, ducks, turkeys, and pigeons. As stated above, the use of bird feeders with a design limiting fecal accumulation, avoiding the production of dust from dried bird feces when cleaning contaminated surfaces, and washing hands afterwards should reduce the risk of contracting psittacosis.

Answer 6.7: The authors of this study list a number of factors that could influence conclusions drawn from their study:

1. This was a cross-sectional survey lacking a time dimension, making it difficult to establish causality. Measurements of possible causal factors and outcomes occurred at only one point in time.
2. The low response rate (31%), which was reduced further to 23% due to inclusion criteria, may have influenced the representativeness of the sample.
3. There is a lack of agreement among researchers regarding exact terminology for some of the constructs that were investigated.

The authors recommend that a follow-up longitudinal study following cohorts of students over time be conducted, examining scores during didactic and clinical study periods, to assess potential differences in ProQOL scores and the factors associated with their change. They also suggest that modification of the survey instrument may be desirable to improve relevance to non-human care.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

7 Measuring and Communicating Prognoses

7.1 EXPRESSING PROGNOSES

Prognosis is a prediction of the expected outcome of disease with or without treatment. Prognosis is expressed as the probability or likelihood that something will occur in the future. The significance of this probability depends on your point of view. Clinical experience may indicate that the likelihood of improvement following a given treatment regimen is 75%, but from the patient's perspective, it's either 0% or 100%. Practitioners should avoid statements that can be misconstrued as a contract—a definite statement about an outcome. Clients must be apprised of the probabilities of unfavorable, as well as favorable, outcomes. The objective is to avoid expressing prognosis with vagueness when it is unnecessary, and with certainty when it is misleading. Breach of contract and malpractice are bases for lawsuits, but "therapeutic reassurance"—the desire to appear positive while making an explanation or obtaining informed consent—are not (Hannah, 1985). Informed consent is an essential part of any contract and, in veterinary practice, it is vital that the client understand the range of treatment options, estimated costs, and the significance and risks of any procedure that a veterinary practitioner may carry out (Anonymous, 2016).

When communicating a prognosis, the practitioner should strive to supply facts and figures that really help the client make a decision. Specifically, a prognosis should include (1) the variability in the course of disease relative to treatment options, (2) a time reference, (3) risk of treatment-related death (or other untoward reaction), (4) cost, and (5) the nature of the benefit attainable (Crow, 1985).

There are few animal diseases that are documented with this kind of clinically useful information. Instead, evaluations of disease frequently document improvement in tissue morphology, changes in blood chemistries, or physiologic adjustments. Although this information may be useful in understanding the origins and mechanisms of disease, it may lack clinical relevance. Wherever possible, prognoses should be assessed in ways that can be perceived by the patient and its owner.

Clinical experience may indicate that the likelihood of improvement following a given treatment regimen is 75%, but from the patient's perspective, it is either 0% or 100%.

EXAMPLE 7.1: HOW IMPORTANT IS THE CHOICE OF ELECTROLYTE SOLUTION IN REHYDRATION THERAPY OF DIARRHEIC NEONATAL CALVES?

Background: Hyperkalemia is a frequently observed electrolyte imbalance in dehydrated neonatal diarrheic calves that can result in skeletal muscle weakness and life-threatening cardiac conduction abnormalities and arrhythmias.

Objectives: Trefz et al. (2017) conducted a prospective, randomized clinical trial to test the hypothesis that intravenous administration of a small-volume hypertonic $NaHCO_3$ (sodium bicarbonate) solution is clinically more effective in decreasing the plasma potassium concentration (cK) in hyperkalemic diarrheic calves than hypertonic NaCl (sodium chloride) or glucose solutions.

Study Design: Randomized controlled clinical trial.

Methods: Twenty-two neonatal diarrheic calves were included in the study. Criteria for inclusion were a clinical diagnosis of neonatal diarrhea, age $\leq$ 21 days, and a measured plasma potassium concentration >5.8 mmol/L. Calves were randomly allocated to one of three treatment groups and received either 8.4% $NaHCO_3$ (n = 7), 7.5% NaCl (n = 8), or 46.2% glucose (n = 7) IV over 5 minutes and were subsequently allowed to suckle 2 L of a commercially available oral electrolyte solution. If the respective volume of the solution was not suckled entirely within 10 minutes, the remainder of the solution was tube-fed after blood sampling at 30 minutes. After the end of the study period at 120 minutes, calves were treated according to clinic principles and received further infusions based on the current acid-base and clinical dehydration status. Infusions with $NaHCO_3$ and NaCl provided an identical sodium load of 6.4 mmol/kg BW. Physical examination followed a standardized protocol and included the clinical assessment of posture/ability to stand, behavior, suckling and palpebral reflex, and extent of enophthalmos before administration of infusion solutions and at the end of the study period at 120 minutes. Scores of clinical variables were expressed as median and corresponding minimum and maximum values and compared for statistically significant differences.

Results: Hypertonic $NaHCO_3$ infusions produced an immediate and sustained decrease in plasma cK. After 120 minutes, the mean decrease in cK from baseline was $-26 \pm 10\%$, $-9 \pm 8\%$, and $-22 \pm 6\%$ in groups $NaHCO_3$, NaCl, and glucose, respectively. After a mean duration of 12 ± 5 days of hospitalization, 21 out of the 22 calves were discharged in a healthy state. Clinical scores for posture, behavior, and degree of enophthalmos improved in all treatment groups, with no difference between treatment groups at the end of the study period. However, a total of 18 calves still showed signs of moderate to severe dehydration as indicated by eye recession into the orbit. Two calves of group NaCl and one calf of group glucose remained unable to stand at the end of the investigation period. One calf of group NaCl had to be euthanatized due to advanced pneumonia, which had progressed during hospitalization.

Conclusions and Significance: Despite the fact that administration of sodium bicarbonate induced the greatest decline in cK among the groups over the 120-minute study period, clinical findings such as posture, behavior, and strength of the suckling reflex were not significantly different among treatment groups. The findings suggest that rehydration should be a primary goal in the treatment of hyperkalemic diarrheic calves.

FOLLOW-UP QUESTION 7.1

What aspects of the study design might explain the contradictory results reported by the authors of this study? (*Hint:* see Chapter 8, "Design and Evaluation of Clinical Trials for a clue.") See Answer 7.1 at the end of this chapter.

Wherever possible, prognoses should be assessed in ways that can be perceived by the patient and its owner.

7.2 NATURAL HISTORY VERSUS CLINICAL COURSE

The **natural history of a disease** describes its evolution without medical intervention. Because of the availability of veterinary services, it is often difficult to obtain information on the natural history of a disease. Once disease is recognized, it is likely to be treated. The **clinical course** of a disease describes its progression once it has come under medical care.

The true natural history of unselected cases of a disease, and the course of those that are recognized, can be quite different. The recognized cases may be a biased sample of all manifestations of the disease that may be particularly symptomatic or may have come to attention because the patients had other symptoms that were not related to the disease. Reports of prognosis from veterinary teaching hospitals and other referral centers may not be representative of cases seen in the typical private practice. Reported cases are often those that had been referred because they were doing poorly.

Reports of prognosis from veterinary teaching hospitals and other referral centers may not be representative of cases seen in the typical private practice.

Knowledge of the natural history of a disease can be useful when counseling animal owners on the prognosis for a particular disease and in monitoring response to therapeutic agents. For example, a number of therapies have been proposed for canine paroxysmal dyskinesias (PD) in dogs, a disease complex of uncertain etiology characterized by a group of hyperkinetic paroxysmal movement disorders. Lowrie and Garosi (2016) described the natural course of the disease in 59 client-owned dogs (36 Labradors and 23 Jack Russell terriers) with clinically confirmed PD that received no medication. Median follow-up time for all dogs was 7 years (range 3 years to 14 years 10 months). Over that time period, 32% of dogs entered spontaneous remission and an additional 42% experienced improvement (decreased frequency and duration of episodes of PD). This study provided useful prognostic information about canine PD. The authors caution that treatment trials for canine PD should consider the **natural history** of this disease in untreated dogs before misattributing remission to specific treatment effects.

7.3 PROGNOSIS AS A RATE

It is convenient to summarize the course of disease as a rate. Rates commonly used for this purpose include survival, case fatality, response, remission, and recurrence. All are expressions of incidence, e.g., events arising in a cohort of patients over time. Two variables that must be considered in the interpretation of rates are assignment of "zero time" and interval of follow-up.

Most reports of prognosis are really based on case series (see Chapter 6). Zero time may be assigned at any point in the course of disease, such as onset of signs, diagnosis, or treatment. Consequently, the computed rates will reflect the way in which zero time is assigned. Cases should be followed for a sufficient period of time for all events to occur. Any period of follow-up that falls short will lower observed rates relative to true ones.

Rates, such as those listed above, are a relatively simple way of expressing prognosis. However, similar overall rates may cover up important differences in prognosis over the course of a disease. Additional information can be extracted from the same data if they are analyzed over time.

7.4 SURVIVAL ANALYSIS

When interpreting a prognosis, we would like to know the likelihood, on average, that individuals with a given condition will experience an outcome of interest at any point in time. When prognosis is expressed as a summary rate, it does not contain this information. However, **survival analysis** provides information about average time to event at any point over the period being monitored. Population models and life table analyses are two techniques commonly used for this purpose.

Similar overall rates may cover up important differences in prognosis over the course of a disease.

7.4.1 Population Models

When populations are in a **steady state**, i.e., constant rates of migration in and out of the population, then the influence of secondary factors such as changes in population dynamics, immunizations, or exposure to infectious agents can be evaluated.

EXAMPLE 7.2: HOW EFFECTIVE IS IMMUNIZATION OF "FREE-RANGE" CHICKENS FOR THE CONTROL OF AVIAN INFLUENZA?

Background: As of 2016, 63 countries have reported circulation of highly pathogenic avian influenza (HPAI) subtype H5N1 (HPAI H5N1) among poultry, and 16 have reported human cases (Villanueva-Cabezas et al., 2017). Indonesia is one of the most severely affected, reporting enzootic circulation of the virus among poultry and the world's second highest number of reported human cases, after Egypt. Control of HPAI H5N1 among poultry in Indonesia is a complex public health challenge, as over 60% of households keep *village chickens*, a backyard production system operated as a subsistence or side business. These production systems generally lack biosecurity measures and may serve as a reservoir of HPAI H5N1 that sporadically infects the commercial sector and human population. Previous disease modeling showed that the rapid turnover of chicken populations might undermine **herd immunity** after vaccination, although actual details of how this effect applies to Indonesia's village chicken population have not been determined (herd immunity is discussed further in Chapter 12, "Establishing Cause").

Objectives: Villanueva-Cabezas et al. (2017) explored the turnover effect in Indonesia's scavenging and mixed populations of village chickens upon the achievable level of herd immunity after vaccination in the Java region of Indonesia.

Study Design: Mathematical (simulation) modeling based on published demographic and production data.

Methods: A steady-state population model was developed driven by data collected from village chicken flocks. The scavenging and mixed populations were divided into four gender-specific age stages: eggs, chicks (<1 month), growers (1–6 months), and adults (>6 months). Survivorship for eggs and chicks was calculated based on published reports on the abundance of individuals at different times of each particular age stage. Given the lack of information on overall survivorship for growers and adult village chickens in the Indonesian context, these were estimated based on published reports on the average flock structure of village chickens in Java. The maximum age for adults for the scavenging chicken population was set at 2 years for hens and 1.5 years for cocks. The investigators assumed a maximum age of 1 year for hens and cocks in the mixed population. Population dynamics were simulated for 208 weeks until a steady state was reached, and then the population turnover effect was simulated for 16 weeks after vaccination in two "best-case" scenarios, where the whole population (scenario 1), or birds aged over 14 days (scenario 2), were vaccinated with 100% immune response.

Results: The investigators found that although the scavenging and mixed populations have different productive traits, both populations are dominated by females, "growers" and "chicks." The simulations predicted that protective herd immunity would last about 2 months for the scavenging, and between 2 and 2.5 months for the mixed populations, depending on the definition of the "critical threshold" (Figure 7.1).

Conclusions and Significance: The authors conclude that Indonesia's village chicken population does not have a unique underlying population dynamic and, therefore, different turnover effects on herd immunity may be expected after vaccination. Nonetheless, simulations

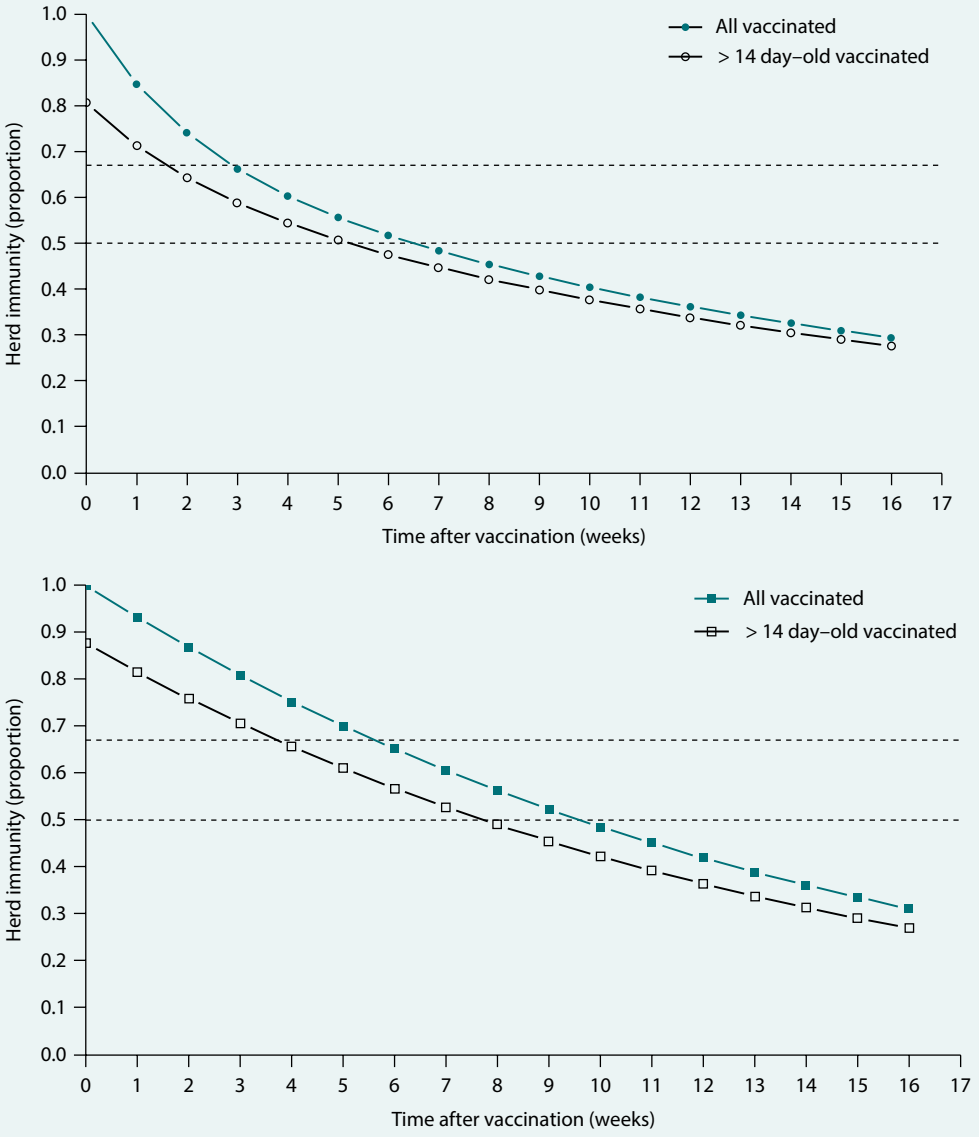

FIGURE 7.1 Herd immunity after a perfect vaccination of a population of scavenging (top) or mixed village chickens (bottom). Dashed lines represent the higher and lower critical herd immunity thresholds that would prevent H5N1 re-emergence. (From Villanueva-Cabezas JP et al. *Zoonoses Public Health* 2017;64:53–62. With permission.)

carried out in best-case scenarios highlight the limitations of current vaccine technologies to control HPAI H5N1 and suggest that additional strategies must be explored.

FOLLOW-UP QUESTION 7.2

For modeling purposes, the investigators divided the scavenging and mixed chicken populations up into the following subgroups: chicks (female and male), female growers, male growers, hens, and cocks. Which subgroup(s) are likely to be the primary factor driving the accelerated loss of herd immunity in these flocks?

7.4.2 Cross-Sectional Studies

Veterinary medicine lacks the kinds of vital statistics data regularly collected for human populations around the world. As a result, epidemiologists must "mine" comparable data from alternative sources, such as questionnaire surveys, patient records, and pet insurance databases. As no reporting requirement exists for animal populations, there is always a potential for selection bias in these datasets. Sampling in these studies is typically cross-sectional, meaning that data are harvested across the entire population over a short period of time, typically a year or two. Death rates for each age group can be used to create a life table (see below), survivorship curve, estimate life expectancy, and evaluate the contribution of various risk factors and pathologic states to outcomes. Since a rate is used, rather than absolute numbers of deaths (dangling numerators), the resulting survival data are unaffected by the number of individuals in each age class.

7.4.2.1 Analysis of Longevity

The following cross-sectional study is a straightforward approach to evaluating canine longevity based on age at death. As only lifespan data were collected, it was not possible to perform a life table analysis based on death rates for each age group. However, the findings are useful for estimating longevity of different breeds of dogs and the relative importance of several risk factors and health status to mortality.

EXAMPLE 7.3: WHAT CAN PATIENT RECORDS TELL US ABOUT CANINE LONGEVITY?

Background: Improved understanding of longevity represents a significant welfare opportunity for the domestic dog, given its unparalleled morphological diversity. Epidemiological research using electronic patient records (EPRs) collected from primary veterinary practices overcomes many inherent limitations of referral clinic, owner questionnaire, and pet insurance data.

Objectives: O'Neill et al. (2013) sought to use a database of merged EPRs from primary veterinary practices in England to quantify canine longevity, establish the most common causes of mortality, and evaluate associations between demographic risk factors and longevity. It was hypothesized that crossbred longevity would exceed that for purebreds, independent of bodyweight.

Study Design: Cross-sectional.

Methods: Clinical health data were retrieved from 102,609 owned dogs attending first opinion veterinary practices (n = 86) between January 2009 and December 2011 in central and southeast England. Data on 5095 confirmed deaths were used to quantify canine longevity, establish the most common causes of mortality, and evaluate associations between demographic risk factors and longevity.

Results: Of deceased dogs for which there was information on breed, sex, and insurance status, 3961 (77.9%) were purebred, 1082 (21.3%) were intact females, 1304 (25.7%) were neutered females, 1464 (28.9%) were intact males, 1224 (24.1%) were neutered males, and 1105 (21.7%) were insured. Overall and breed-specific (for breeds with 20 or more study dogs) longevities were reported in terms of median, interquartile range (IQR), and range. Overall longevity was bi-modally distributed, peaking in years 1 and 14, with similar distribution patterns for purebred and crossbred dogs (Figure 7.2). The overall median longevity was 12.0 years (IQR 8.9–14.2). The longest-lived breeds were the Miniature poodle (14.2 years), Bearded collie (13.7 years), Border collie (13.5 years), and Miniature dachshund (13.5 years), while the shortest-lived were the Dogue de Bordeaux (5.5 years) and Great Dane (6.0 years). The most frequently attributed causes of death were neoplastic, musculoskeletal, and neurological disorders. The results of multivariable modeling indicated that

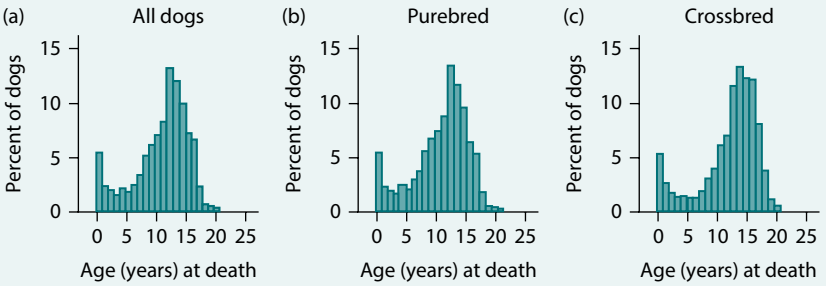

FIGURE 7.2 Distribution patterns for age at death of dogs attending primary veterinary practices in England, showing the percentage that died within 1-year age bands. (a) All dog types (n = 5095); (b) purebred dogs (n = 3961); (c) crossbred dogs (n = 1124). Note: Ten records held no breed data. (From O'Neill DG et al. *Vet J* 2013;198:638–643. With permission.)

longevity in crossbred dogs exceeded purebred dogs by 1.2 years (95% confidence interval 0.9–1.4; $p < 0.001$) and that increasing bodyweight was negatively correlated with longevity.

Conclusions and Significance: The current findings highlight major breed differences for longevity and support the concept of hybrid vigor in dogs.

FOLLOW-UP QUESTION 7.3

What are possible sources of bias/confounding in this study?

7.4.2.2 Life Table Analysis

Example 7.3 estimated canine longevity based on age at death of patients when under veterinary care. A more precise estimate of canine longevity and life expectancy can be obtained through **life table analysis**. There are two principal forms of the life table: the cohort (or generation) life table and the current life table. The **cohort (or generation) life table** records the actual mortality experience of a particular group of individuals (the cohort) over its entire lifetime. The **current (or period) life table** gives a cross-sectional view of the mortality and survival experience of a population during a current year and is dependent on the age-specific death rates prevailing in the year for which it is constructed. The current life table is the most effective means of summarizing the mortality and survival experience of a population and forms a sound basis for statistical inferences about the population under study (Chiang, 1979).

The following study used data from dogs enrolled in a health insurance program to create a current life table that was used to evaluate causes of death and life expectancies.

EXAMPLE 7.4: HOW CAN PET INSURANCE DATA BE USED TO PREDICT CANINE LIFE EXPECTANCY?

Background: Knowing the life expectancy of a defined animal population and the risk factors that may affect it can provide information required to assist in the prevention and control of disease and the education of owners.

Objectives: Inoue et al. (2015) used data from a Japanese veterinary care insurance company to construct a life table of dogs in Japan and to determine common causes of death.

Study Design: Case series.

Methods: Data for life table analysis were collected from a total of 299,555 dogs for 1 year following entry into, or renewal of, an insurance policy any time during fiscal year 2010 (April 1, 2010–March 31, 2011). In constructing the life table, the authors used an age interval of 1 year. The probability of a dog dying in the age interval (x, x + 1) was calculated by dividing the number of dogs that died during that age interval by the number of dogs alive at the beginning of that interval (x). The fraction of last year of life for age x was calculated as the average of the fraction of the last year of life for dogs that had died during the interval (x, x + 1). These parameters were used to construct life tables for all breeds combined and for five groups of breeds grouped according to their ideal body weights in accordance with the method described by Chiang (1979) using an initial hypothetical population of 10,000 dogs (Table 7.1). The life expectancy at age x was calculated as the number of years, on average, yet to be lived by a dog of age x. The causes of death for the 4169 dogs that left the insurance program due to death were categorized into one of the 18 diagnostic categories by body system or type of disease used by veterinarians when they complete the insurance claim form for the owners.

Results: The overall life expectancy of dogs was 13.7 years. Life expectancies for different breed groups during the first 10 years of life are depicted in Figure 7.3. The probability of death was high in the first year of life, lowest in the second and third years, and increased exponentially after 3 years of age. As body weight increased, life expectancy tended to decrease, except in the 5–10 kg body weight group. The life expectancy was 13.8 years in

TABLE 7.1
Current Life Table of Insured Dogs in Japan (All Breeds Combined)

Age Interval in Years	Probability of Dying in Interval (x, x + 1)	Number Living at Age x	Number Dying in Interval (x, x + 1)	Fraction of Last Year of Life	Number of Years Lived in Interval (x, x + 1)	Total Number of Years Lived Beyond Age x	Expectation of Life at Age x
x to x + 1	$\hat{q}_x$	l_x	d_x	a'_x	L_x	T_x	$\hat{e}_x$
0–1	0.0107	10,000	107	0.48	9945	136,580	13.7
1–2	0.0049	9893	48	0.45	9867	126,636	12.8
2–3	0.0041	9845	41	0.48	9824	116,769	11.9
3–4	0.0048	9804	47	0.54	9783	106,945	10.9
4–5	0.0066	9757	64	0.51	9726	97,163	10.0
5–6	0.0082	9693	79	0.51	9654	87,437	9.0
6–7	0.0102	9614	98	0.53	9568	77,783	8.1
7–8	0.0152	9516	145	0.51	9445	68,215	7.2
8–9	0.0254	9371	238	0.49	9250	58,770	6.3
9–10	0.0388	9133	354	0.49	8952	49,520	5.4
10–11	0.0606	8778	532	0.50	8512	40,568	4.6
11–12	0.0877	8246	723	0.51	7889	32,056	3.9
12–13	0.1266	7523	953	0.48	7024	24,167	3.2
13–14	0.1688	6570	1109	0.45	5961	17,143	2.6
14–15	0.2678	5461	1463	0.49	4708	11,181	2.0
15–16	0.3739	3999	1495	0.45	3178	6473	1.6
16–17	0.4715	2504	1181	0.38	1777	3295	1.3
17+	1.0000	1323	1323	–	1518	1518	1.1

Source: Inoue M et al. *Prev Vet Med* 2015;120:210–218. With permission.

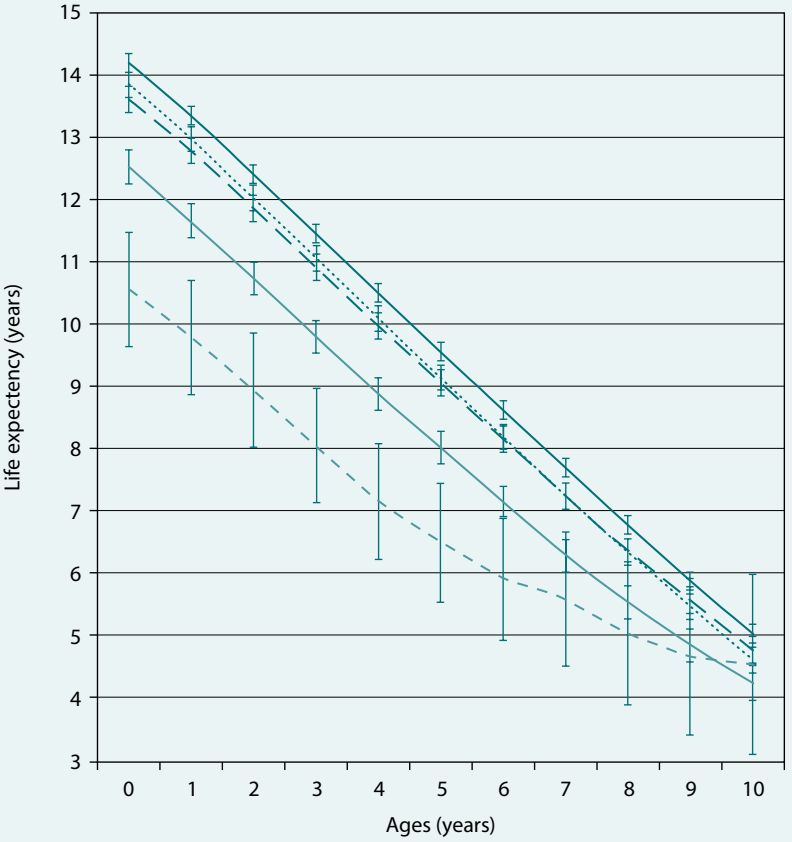

FIGURE 7.3 Life expectancy up to 10 years of age of toy, small, medium, large, and giant breeds of insured dogs in Japan. Dotted, solid, dashed, gray, and gray dashed lines indicate toy, small, medium, and large breed groups, respectively. The error bars indicate 95% confidence intervals. (From Inoue M et al. *Prev Vet Med* 2015;120:210–218. With permission.)

the <5 kg body weight group, 14.2 years in the 5–10 kg body weight group, 13.6 years in the 10–20 kg body weight group, 12.5 years in the 20–40 kg body weight group, and 10.6 years in the ≥40 kg body weight group. The probability of death increased as dogs got older for most potential causes of death. Neoplasia resulted in the highest probability of death, especially in the large and giant breed groups. Cardiovascular system disorders were the second major cause of death, and the toy group had a probability of death significantly higher than the other breed groups at age 12+.

Conclusions and Significance: This study identified risk factors that may affect the longevity of dogs in Japan. The findings could be used by pet owners, veterinary clinicians, and breeders to introduce measures promoting the health care of dogs in general and of certain breeds at different ages in particular.

FOLLOW-UP QUESTION 7.4

Example 7.3 estimated longevity, whereas Example 7.4 estimated life expectancy. What methodological approach used in Example 7.4 accounts for this difference?

7.4.3 Longitudinal Studies

The most direct way of learning about survival is to assemble a cohort of individuals with the condition of interest and periodically count the number remaining throughout the course of their illness. Cohort studies evaluate outcomes over time for one or more groups that have something in common. Clinical trials are cohort studies that describe the prognosis for at-risk or diseased patients with or without treatment. The results may be expressed in life tables and as survival curves.

Maintaining the integrity of a cohort is often difficult because (1) participants ordinarily become available for a study over a period of time, thus resulting in variable duration of follow-up, and (2) they may drop out of the study before the end of the follow-up period. **Censoring** can be used to more efficiently use follow-up data, regardless of the time at which an individual enters or leaves a study, by only using data up to the point of the subject's last follow-up. Thereafter they are dropped, or **censored**, from the population at risk. Results can be expressed in fixed chronological intervals or as **time-to-event intervals** (Kaplan–Meier analysis).

In the **Kaplan–Meier method**, each time interval is defined by the occurrence of the outcome of interest (the "event"). In many studies this is the death of a patient. The probability of surviving over each time interval is calculated by dividing the number of patients surviving by the number at risk of dying over that interval. Individuals who have already died, dropped out of the study, or have not been followed up to that point are not included in the population at risk for that interval. The probability of surviving to any point in time since the beginning of the study ("0 time") is obtained by multiplying the probability of surviving over the preceding time interval by the probability of surviving up to the beginning of that interval. The mechanics of these calculations are illustrated in the following example that applies Kaplan–Meier analysis to data from a cohort study of feline survival following splenectomy.

EXAMPLE 7.5: WHAT IS THE PROGNOSIS FOR CATS UNDERGOING SPLENECTOMY?

Background: Complete splenectomy is indicated to treat suspected malignant neoplasia, ischemic obstruction, or generalized splenic enlargement secondary to infiltrative diseases of the spleen. It may also be used as part of the treatment for some immune-mediated diseases. The types, treatments, and prognoses of splenic diseases in cats are not well described. The most common splenic diseases in cats include mast cell tumors (MCTs), lymphoma, and myeloproliferative disease. No studies have yet examined whether there are prognostic factors for survival of cats undergoing splenectomies.

Objectives: Gordon et al. (2010) conducted a study to identify common reasons for and clinical predictors of survival for cats undergoing splenectomy.

Study Design: Case series.

Methods: Medical records of 19 cats that had complete splenectomies between July 1999 and July 2005 were reviewed to identify preoperative, intraoperative, and postoperative factors that might affect the survival of cats undergoing splenectomy. Follow-up data were collected through record review, veterinarian contact, or owner contact.

Results: The most common presenting signs were a palpable abdominal mass in 58% and anorexia in 47% of the cats. Mast cell tumors were the most common reason for splenectomy and were found in 10/19 cats (53%), followed by hemangiosarcoma in 4/19 (21%), and lymphoma in 2/19 (11%). The Kaplan–Meier median survival time (MST) was 197 days, with a range from 2 days to 1959 days. Three cats were noted to have preoperative weight loss, and this was the only factor that had prognostic significance for survival following surgery. For

cats with weight loss, the MST was 3 days compared with 293 days for cats with no weight loss noted ($p = 0.008$). Thirteen cats died over the 1959-day postoperative follow-up period, while six cats were still alive from 339 to 1959 days post-surgery. The original data appear in Table 7.2. Survival analysis of these data is complicated by **censored observations**—six cats that did not die but were lost to follow-up over the course of the study. A Kaplan–Meier life table analysis of the data, which adjusts for the censored observations, is depicted in Table 7.3, and the corresponding survival curve in Figure 7.4. The population at risk over each interval is adjusted for previous deaths and loss to follow-up. Thus, even though surviving cats were not followed for the same period of time, each contributed to the analysis for the period that it remained in the study.

Conclusions and Significance: The most common type of feline splenic disease resulting in splenectomy was MCT. The only prognostic factor for survival time following a splenectomy in cats that was identified in this study was presence of preoperative weight loss.

TABLE 7.2

Original Survival Data for All Cats Undergoing Splenectomies

Group	Time to Event (Days)
Still alive at last follow-up	339, 343, 854, 1270, 1789, 1959
Died during follow-up	2, 2, 3, 3, 4, 25, 53, 132, 197, 259, 293, 322, 415

Source: Gordon SS et al. *J Feline Med Surg* 2010;12:256–261.

TABLE 7.3

Kaplan–Meier Life Table Analysis of Data from Table 7.2 on Feline Splenectomies

Interval (Days)	No. of Events			Survival	
	Censored	Death	At Risk	Interval (%)	Overall (%)
0	0	0	19	–	100
2	0	2	19	89	89
3	0	2	17	88	79
4	0	1	15	93	74
25	0	1	14	93	68
53	0	1	13	92	63
132	0	1	12	92	58
197	0	1	11	91	53
259	0	1	10	90	47
293	0	1	9	89	42
322	0	1	8	88	37
415	2	1	5	80	29
854	1	0	4	100	29
1270	1	0	3	100	29
1789	1	0	2	100	29
1959	1	0	1	100	29

Source: Gordon SS et al. *J Feline Med Surg* 2010;12:256–261.

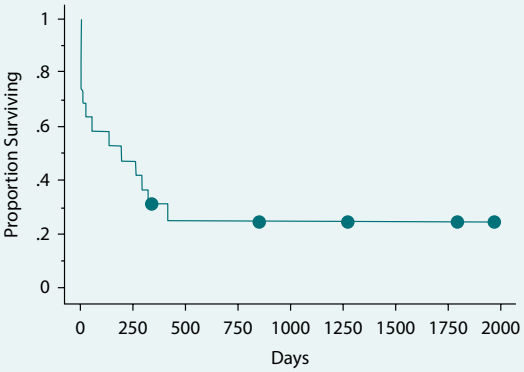

FIGURE 7.4 Kaplan–Meier survival curve for all cats undergoing splenectomies. Censored observations are depicted as dots. Two cats are represented by the first dot corresponding to days 339 and 343. See Table 7.3 for Kaplan–Meier source data. (From Gordon SS et al. *J Feline Med Surg* 2010;12:256–261. With permission.)

FOLLOW-UP QUESTION 7.5

The authors report that median survival time derived from the Kaplan–Meier life table analysis was 197 days. Why is the median preferred over the mean (average) to express survival?

Life table analysis and survival curves can be used to describe outcomes of disease besides death, e.g., recurrence of a tumor, remission duration, rejection of graft, or reinfection, and to identify prognostic factors for these outcomes. Clinical and pathologic staging criteria are often correlated with patient outcomes expressed as survival curves. In fact, the frequency of any event can be studied by means of life tables, so long as the outcome is dichotomous (i.e., either/or) and can occur only once during the follow-up period.

7.4.4 Interpreting Survival Curves

Several points must be considered when interpreting survival curves. First, since the data include censored observations, the percentage of individuals at each data point may not be equivalent to the actual number of individuals remaining in the study. This can be appreciated through an examination of Table 7.3 and Figure 7.4. The last five data points (415–1959 days) each represent 29% (5.5) of 19 cats present at the beginning of the study. The actual number of cats over this period declines from 5 to 1 as they are lost to follow-up.

Second, the number of individuals at risk declines as we move from left to right along the survival curve (Figure 7.4). Consequently, our estimates of the probability of survival depend on what happens to fewer and fewer individuals. A single event toward the end of the follow-up period will have a much greater impact than at the beginning. As a result, we can have less confidence in our estimates of survival toward the end of the survival curve.

Finally, the survival curve reflects the effect of a survival rate upon a steadily decreasing population at risk. This accounts for the steadily decreasing slope of the survival curve over the follow-up period. Although the percentage survival may appear to improve over time, the survival rate may actually remain unchanged. This is similar to a radioactive decay curve whose shape reflects the steady decay of a radionuclide over time.

7.5 COMMUNICATION OF PROGNOSES

The use of qualitative terms to express chances of success or failure is inherently ambiguous. Furthermore, veterinarians frequently do not agree regarding the prognosis for many common

illnesses. Unfortunately for veterinary clinicians, there is no definitive source of prognostic information about diseases of domestic animals. Furthermore, there may be a mismatch between what the veterinarian is communicating and what the client understands. The definition of recovery, for example, can be a source of misunderstanding. Attention to communication competency should be an important part of the training of veterinary students, practitioners, and support staff.

REFERENCES

Anonymous. Informed consent: What does it mean and how is it achieved? *Vet Rec* 2016;178:463.

Chiang, CL and World Health Organization. *Life Table and Mortality Analysis/Chin Long Chiang*. Geneva: World Health Organization; 1979. Full text available at: http://www.who.int/iris/handle/10665/62916.

Crow SE. Usefulness of prognoses: Qualitative terms vs quantitative designations. *J Am Vet Med Assoc* 1985;187:700–703.

Gordon SS, McClaran JK, Bergman PJ et al. Outcome following splenectomy in cats. *J Feline Med Surg* 2010;12:256–261.

Hannah HW. Promising a result. *J Am Vet Med Assoc* 1985;186:1166.

Inoue M, Hasegawa A, Hosoi Y et al. A current life table and causes of death for insured dogs in Japan. *Prev Vet Med* 2015;120:210–218.

Lowrie M and Garosi L. Natural history of canine paroxysmal movement disorders in Labrador retrievers and Jack Russell terriers. *Vet J*. 2016;213:33–37.

O'Neill DG, Church DB, McGreevy PD et al. Longevity and mortality of owned dogs in England. *Vet J* 2013;198:638–643.

Trefz FM, Constable PD, and Lorenz I. Effect of intravenous small-volume hypertonic sodium bicarbonate, sodium chloride, and glucose solutions in decreasing plasma potassium concentration in hyperkalemic neonatal calves with diarrhea. *J Vet Intern Med* 2017;31:907–921. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5435070/

Villanueva-Cabezas JP, Campbell PT, McCaw JM et al. Turnover of village chickens undermines vaccine coverage to control HPAI H5N1. *Zoonoses Public Health* 2017;64:53–62.

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 7.1: The authors sought to demonstrate the efficacy of an intravenous sodium bicarbonate solution for reducing plasma potassium concentration (cK) in diarrheic neonatal calves. However, neither clinical findings nor discharge status were significantly different among treatment groups. Clinical recovery may have in fact been due to the beneficial effects of concomitant hydration therapy administered to all calves. The study design did not permit separate evaluation of this treatment strategy. Further, it is not clear whether the investigators were blinded to treatment status when assessing clinical status shortly after electrolyte therapy. The authors emphasize that the aim of the study was to compare the potassium-lowering efficiency of different hypertonic infusion solutions in the initial treatment of hyperkalemic calves and not to test the resuscitative effect (clinical relevance) of a single injection of hypertonic infusion solution and subsequent suckling of an oral electrolyte solution.

Answer 7.2: The authors conclude that fecundity of hens, the survival of eggs that become chicks, and the survival of female growers becoming hens have the largest impact on the reduction in herd immunity. New individuals (hatchlings) coming into the population were susceptible to influenza. The population model predicted that 44% of the mixed population and 55% of the scavenging population would be replaced after 60 days. Thus, the effectiveness of the vaccination program is affected more by population dynamics than by the efficacy of the vaccine, suggesting that to achieve sustained impact, immunization of flocks would need to be conducted at least every second month, which may not be feasible in the Indonesian context.

Answer 7.3: The data gathered were essentially a "snapshot" of each patient, e.g., a cross-sectional study design. Patients were not followed longitudinally but rather patient histories were constructed retrospectively. A total of 5095 patients were studied, but outcomes for the remaining 97,514 patients (95% of the clinic population) are unknown. Because of the nature of the patient population (under

veterinary care), this is not really a study of life longevity. Only practice-attending dogs were included, so data were not captured for dogs not under veterinary care. It was possible that death data were not captured for some dogs that died at home or at emergency out-of-hours clinics. The results for neutering association should be interpreted with caution because this variable was modeled as time independent (i.e. a single value applied throughout life), but lifetime ovary exposure and longevity depend upon the age that neutering was performed. There were also missing data in the study population. The cause of death was known for only 87%, and adult bodyweight was available for only 50.3% of study subjects.

Answer 7.4: The difference between Example 7.3 and 7.4 is that rather than simply calculating the lifespan of dogs based on age at death (Example 7.3), age-specific death rates were calculated and used to create a life table that reflected the mortality and survival experience of the population during each age interval, and life expectancy from that point forward. Further, data from the entire population were used in Example 7.4, not just from animals that had died as in Example 7.3.

Answer 7.5: The median is the middle value or "halfway point" where half of the cats have died and half are still alive. A splenectomized cat in this study had a 50:50 chance of surviving 197 days. The median is unaffected by extreme values as found in skewed data distributions such as this one. The mean survival time for this dataset was 435 days, but would be an inappropriate representation of survival as 6 of 19 cats were still alive at their last follow-up and, according to the life table, only 29% of cats survived 435 days.

8 Design and Evaluation of Clinical Trials

8.1 INTRODUCTION

Throughout this text, a distinction has been made between epidemiologic studies of naturally occurring disease and laboratory studies of experimentally induced disease. Within the field of clinical epidemiology, the evaluation of treatment effects (the clinical trial) comes as close to a laboratory experiment as any activity that we have discussed. In evaluating clinical trials, the practitioner must consider not only whether the data support the author(s)' conclusions, but also whether the study design was appropriate for the question being asked. In this chapter, we first examine factors that can influence the outcome of clinical trials and then apply these criteria to selected case studies.

> ***Treatments should be adopted not because they ought to work, but because they do work.***

Therapeutic hypotheses may come from an understanding of the mechanisms of disease, clinical observations, or epidemiologic studies of populations. Regardless of their source, new treatment regimens must be tested. In other words, treatments should be adopted not because they ought to work, but because they do work (Fletcher et al., 1982).

8.2 EFFICACY, EFFECTIVENESS, AND COMPLIANCE

Efficacy is a measure of how well a treatment works among those who receive it. **Effectiveness**, on the other hand, is a measure of how well a treatment works among those to whom it is offered. **Compliance** is a measure of the proportion of patients (or their owners) that adhere to a prescribed treatment regimen. Thus, an efficacious treatment could be ineffective due to poor compliance. This relationship can be summarized as

$$\text{Effectiveness} \approx \text{Efficacy} \times \text{Compliance}$$

Intention-to-treat (ITT) analysis considers the outcome for all subjects entered into a trial, regardless of whether they received the treatment they were actually supposed to receive, e.g., analysis according to treatment assigned rather than treatment received. It is a measure of treatment effectiveness. It addresses the question actually faced by clinicians: Which treatment choice is best at the time the decision must be made? **Per-protocol** analysis only considers the outcome for subjects that actually received an intervention, regardless of the group to which they were originally assigned (Fletcher et al., 2014). ITT analyses may prevent overestimation of treatment efficacy in case of substantial withdrawal of study subjects, as in response to adverse drug effects (Olivry and Mueller, 2003).

8.3 CLINICAL TRIALS: STRUCTURE AND EVALUATION

Practitioners initiate an observational study of treatment effects every time they treat a patient. However, because of the many potential sources of bias during routine patient care, a more formal approach to evaluating treatment outcomes is usually required. The **clinical trial** is a cohort study

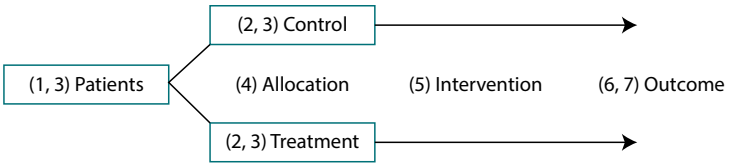

FIGURE 8.1 Design and potential sources of bias (see Table 8.1) in clinical trials. (From Fletcher RH et al. *Clinical Epidemiology—The Essentials*. Baltimore: Williams and Wilkins; 1982. With permission.)

specifically designed to facilitate the detection and measurement of treatment effects, free of extraneous variables. Because of the experimental nature of clinical trials, they are sometimes referred to as **intervention or experimental studies**.

The design and potential sources of bias in a clinical trial are depicted in Figure 8.1 and summarized in Table 8.1. They are discussed in greater detail in Sections 8.3.1–8.3.8 below. When designing a clinical trial, the first step should be a determination of the minimum number of subjects required to achieve the desired level of **statistical power**. Too few subjects and random variation in outcome may obscure the effects of a beneficial treatment. Study subjects are allocated to either treatment or control groups. Both are treated identically with the exception that the treatment group receives an intervention that is believed to be beneficial. The control group usually receives a **placebo**, an intervention designed to simulate the act of treatment but lacking its beneficial component(s). Any differences that emerge between the two groups over time are attributed to the treatment. Virtually any parameter can be used to measure and express the outcome of a clinical trial. In veterinary medicine, the outcome may be expressed in terms of the health benefit to the patient, or as productivity or economic benefits.

The clinical trial is a cohort study specifically designed to facilitate the measurement of treatment effects, free of extraneous variables.

There are two measures of validity for clinical trials: internal and external. **Internal validity** refers is the extent to which conclusions drawn from a study are correct for the sample of patients being studied. **External validity** (**generalizability**) is the degree to which results of a study can be generalized to the population at large from which the sample was drawn, e.g., the **target population**. The first requirement for external validity is internal validity, e.g., invalid conclusions from a clinical trial will also be invalid when applied to the broader population of patients. However, a study may produce valid results but still lack external validity because study subjects are not representative of the general patient population. Examples might be clinical trials whose patient composition does not accurately reflect the gender, age distribution, or clinical severity of patients at large. External

TABLE 8.1
Factors That May Influence the Outcome and Relevance of Clinical Trials

1. Is the case definition explicit, exclusive, and uniform?
2. Is a comparison group explicitly identified?
3. Are both treated and control patients selected from the same time and place?
4. Are patients allocated to treatment and control groups without bias?
5. Is the intended intervention, and only that intervention, experienced by all of the patients in the treated group and not in the control group?
6. Is the outcome assessed without regard to treatment status?
7. Is the method used to determine the significance of the observed results defined explicitly? Can we be certain that the observed results could not have occurred by chance alone?

validity can be maximized by selecting study subjects that are as similar as possible to the patient population to which the results are to be generalized.

Many factors (biases) can affect the internal validity of cohort studies of risk, prognosis, and treatment (Sackett, 1979). These generally originate from one of the following sources (Fletcher et al., 2014):

1. *Selection bias*: Selection (or assembly) bias occurs when the criteria for inclusion of patients in a study do not ensure uniformity of individuals. Patients may differ in ways that are not under study and that can affect the outcome.
2. *Measurement bias:* Measurement bias occurs when uniform standards for measurement of clinical events cannot be maintained over time.
3. *Confounding*: Confounding occurs when two factors are associated with each other, or "travel together," and the effect of one is confused with or distorted by the effect of the other. The effect of confounding is usually dealt with during data analysis, after the study is over.

The criteria outlined in Table 8.1 have proven useful for reducing bias in cohort studies. The points at which they influence the outcome of a clinical trial are indicated in Figure 8.1 and discussed in greater detail below.

Many factors can affect the outcome of cohort studies of risk, prognosis, and treatment. These generally originate from assembly, migration, measurement, or confounding bias.

8.3.1 Case Definition

The first step in a clinical trial is selection of patients who meet the **case definition**. This is not as easy as it might first appear. It may be difficult to define a set of disease signs that will include all true cases of a disease and exclude similar, but unrelated, conditions. Few cases will show the complete range of disease signs and symptoms; thus, minimal criteria for a diagnosis often have to be established. As the number of signs and symptoms required to meet the case definition increases, the definition becomes more and more restrictive and includes a progressively smaller number of cases. Furthermore, the criteria used for the case definition should be uniformly applied when multiple clinics are involved. **Misclassification bias** (a form of **information bias**) occurs when the assignment of subjects to groups (such as cases or controls, or exposure status) are erroneous. This may result, for example, from limited sensitivity and/or specificity of a diagnostic test, or from inadequacy of information derived from medical or other records (Gordis, 2014).

8.3.2 Uncontrolled Clinical Trials

In **uncontrolled clinical trials**, the effects of treatment are assessed by comparing patients' clinical courses before and after treatment, without reference to an untreated comparison group, to see whether an intervention changes the established course of disease in individual patients. The difficulty in interpreting the results of an uncontrolled trial relates to the predictability of the course of disease.

For some conditions, the prognosis without treatment is so predictable that an untreated control group is either unnecessary or unethical. In most cases, however, the clinical course is not so predictable. Some diseases normally improve after an initial attack. If a treatment is given at this time, it may be mistakenly credited with the favorable outcome. Clients tend to seek care for their animals when signs are at their worst. Patients sometimes begin to recover after seeing the veterinarian because of the natural course of events (natural history of the disease), regardless of what was done. Severe diseases that normally are not self-limiting may nonetheless undergo spontaneous remission. In these cases, improvement in the patient's condition would mistakenly be attributed to the treatment if it had been initiated when signs were most evident.

EXAMPLE 8.1: WHAT IS THE CLINICAL COURSE OF EQUINE SARCOID IN YOUNG HORSES?

Background: Equine sarcoids (ES) are the most common skin neoplasia in equids, accounting for up to 90% of all cutaneous tumors. Sarcoids invade dermal and/or subcutaneous tissues locally, but true metastatic dissemination does not occur. Two common forms of ES are occult (roughly circular hairless areas of skin) and verrucous (wart-like). Although the disease is rarely fatal, tumors may become ulcerated or infected, and recurrence is frequently observed after tumor removal. Accordingly, welfare and economic aspects must be considered when treating this disease. The progression of ES is notoriously unpredictable. Making a choice for the appropriate treatment is challenging when dealing with milder manifestations of ES.

Objectives: Berruex et al. (2016) investigated the clinical course of ES in young horses with and without therapeutic interventions.

Study Design: Non-randomized controlled clinical trial.

Methods: A cohort of 61 ES-affected 3-year-old Franches-Montagnes horses and a breed-, age-, and geographically matched control group of 75 ES-free peers were examined twice over a period of 5–7 years. Owners and caretakers were queried using a standardized questionnaire.

Results: More than half (38/61 = 62%) of the horses that were ES-affected at the age of 3 had become ES free at the time of follow-up (age 8–11). In 29 of 38 horses, representing 48% of the entire ES study population, lesions had spontaneously disappeared without therapy. At the time of follow-up examination, 6 (8%) of the 75 horses of the control group had acquired ES lesions. Of 12 horses that received specific treatment for ES disease, therapy was successful in eliminating the ES lesions in half of them. When differentiating the clinical types of ES lesions, occult ES underwent complete spontaneous regression in 65% (11/17), while verrucous lesions regressed spontaneously in 32% (9/28). None of the evaluated intrinsic or environmental factors showed a significant effect on the risk for development, regression, or exacerbation of ES disease.

Conclusions and Significance: The results document a surprisingly high rate of spontaneous ES regression for young horses affected with milder manifestations of ES disease. These findings justify a "wait-and-see" approach in selected cases of occult and verrucous ES, provided that all lesions are closely monitored. Furthermore, results of this study should also be considered when critically assessing treatment effects of therapies directed against ES, especially in the context of uncontrolled studies. The results suggest that any therapeutic regimen may yield positive results regardless of efficacy.

FOLLOW-UP QUESTION 8.1

What study design is most vulnerable to misinterpretation of results based on the clinical course of equine sarcoid reported in this study? See Answer 8.2 at the end of this chapter.

8.3.3 Comparisons across Time and Place

Diagnosis and treatment strategies change over time. Similarly, the nature of patients, clinical expertise, and medical procedures differ among clinical settings. Thus, the time and place in which conditions are diagnosed and treated can affect the expected prognosis. Clinical trials in which treatment and comparison groups are selected at the same time (**concurrent controls**) and place are less likely to be biased. However, a historical comparison group (**historical controls**) may be the only alternative when it is ethically inappropriate to withhold a promising new treatment from client-owned animals.

EXAMPLE 8.2: HOW DO NEW TREATMENT MODALITIES FOR RELIEVING URETERAL OBSTRUCTION IN CATS, SUCH AS URETERAL STENTING, COMPARE WITH TRADITIONAL SURGERY?

Background: Ureteral obstruction in cats is uncommon but can lead to life-threatening acid-base and electrolyte disturbances. Affected cats are often examined for nonspecific clinical signs such as vomiting, lethargy, and anorexia, and frequently develop azotemia (abnormally high levels of urea or creatinine in the blood).

Objectives: Culp et al. (2016) compared the outcome for cats with benign ureteral obstructions treated by means of ureteral stenting with that of a historical cohort of cats treated by means of the more traditional ureterotomy (surgical removal of the ureterolith) only.

Study Design: Non-randomized controlled clinical trial.

Methods: Data were recorded prospectively on 26 cats treated with ureteral stenting between 2010 and 2014 and compared with medical records data from 36 cats previously treated with ureterotomy at the same veterinary teaching hospital between 2003 and 2009. Procedural complications included the need for at least 1 ureterotomy in 5 of the 26 (19%) ureteral stenting cats to allow for guidewire passage when the guidewire would not pass a ureterolith.

Results: Cats treated with ureteral stents had significantly greater decreases ($p < 0.05$) in blood urea nitrogen (BUN) and serum creatinine concentrations 1 day after surgery and at hospital discharge compared with values for cats that underwent ureterotomy. Of the 26 cats in the ureteral stenting group, 24 (92%) were discharged versus 28 (78%) of the 36 cats in the ureterotomy group. The magnitude of the response to the two treatments is summarized in Table 8.2.

TABLE 8.2

Results of Serum Biochemical Analyses Performed at the Time of Hospitalization, 1 Day after Ureteral Stenting or Ureterotomy, and at the Time of Hospital Discharge

	Ureteral Stenting		Ureterotomy[a]		Reference	
Variable	No.	Median (Range)	No.	Median (Range)	Range	*p* value[b]
At time of hospitalization						
Creatinine (mg/dL)	26	8.4 (1.6–20.5)	35	6.7 (1.3–26)	1.1–2.2	0.635
BUN (mg/dL)	26	129 (27–308)	35	89 (22–378)	18–33	0.833
1 day after surgery						
Creatinine (mg/dL)	26	4.2 (1.1–17.7)	28	5 (0.8–27.7)	1.1–2.2	0.058
BUN (mg/dL)	26	70 (19–294)	28	71 (15–286)	18–33	0.048
At time of discharge						
Creatinine (mg/dL)	24[c]	2.2 (1.1–7.9)	27	2.6 (1.2–8.5)	1.1–2.2	0.034
BUN (mg/dL)	24[c]	32 (19–294)	27	42 (14–150)	18–33	0.042

Source: Culp WTN et al. *J Am Vet Med Assoc* 2016;249:1292–1300. With permission.

[a] Complete medical records data and laboratory values were not available for all 36 cats enrolled in the ureterotomy group.

[b] The *p* values represent results of a *t* test (normally distributed data) or the Mann Whitney *U* test (non-normally distributed data) comparing variables for the two groups.

[c] One cat developed acute respiratory distress and died 11 days after surgery despite improvements in serum creatinine and BUN concentrations. The other cat was euthanized after several days of worsening azotemia despite an initial improvement in renal parameters within the first 24 hours after surgery.

Conclusions and Significance: Results suggest that cats with benign ureteral obstructions treated with ureteral stenting were more likely to have resolution of azotemia prior to hospital discharge compared with cats undergoing ureterotomy alone. The authors conclude that the results of ureteral stenting were encouraging, but further investigation is warranted.

FOLLOW-UP QUESTION 8.2

Five of the 26 cats treated with ureteral stents required ureterotomies to allow for guidewire passage when the guidewire would not pass a ureterolith. Despite this alteration in protocol, these cats remained in their respective treatment group for analytical purposes. What analytical strategy does this represent? See Answer 8.2 at the end of this chapter.

8.3.4 Allocating Treatment

Selection bias occurs when the way in which subjects are assigned to study groups influences the results, irrespective of whether an actual cause-effect association exists. When concurrent controls are used, assignment of study subjects to treatment or comparison groups can be done in several ways. Some are more prone to selection bias than others.

1. *Non-random allocation*: If the clinician or owner decides how a case is to be treated, then allocation is considered to be non-random. This approach is prone to systematic differences among treatment groups. Many factors, such as severity of illness, concurrent diseases, local preferences, owner cooperation, etc. can affect treatment decisions. As a result, it is difficult to distinguish treatment effects from other prognostic factors when non-random allocation to treatment groups is used.
2. *Random allocation*: The best way to study unique effects of a clinical intervention is through randomized controlled trials in which patients are randomly allocated to treatment and comparison groups. The purpose of randomization is to achieve an equal distribution of all factors related to prognosis among treatment and control groups. If the number of patients is small, the investigator can compare the distribution of a number of patient characteristics among the groups to assure that randomization has been achieved.
3. *Stratified randomization*: If certain patient characteristics are known to be related to prognosis, then patients can first be allocated to groups (strata) of similar prognosis based on this characteristic and then randomized separately within each stratum. Although stratification can be accomplished mathematically after the data are collected, prior stratification reduces the likelihood of unequal cohorts during the randomization process.

8.3.5 Remaining in Assigned Treatment Groups

It is not uncommon for patients in treatment or comparison groups to cross over into another group or drop out of the study entirely. These are forms of **selection bias** (Gordis, 2014), and the way in which these deviations from protocol are handled depends on the question being asked in the clinical trial. **Explanatory trials** are designed to assess the **efficacy** of a treatment. Treatment outcomes are measured only in those patients who actually receive it, regardless of where they were originally assigned. Thus, patients who fail to adhere to the treatment plan or drop out of the study are ignored, and those who transfer into the treatment group may be included. Results are typically subjected to a **per-protocol analysis** to assess outcomes.

Management trials seek to determine how **effective** a treatment is among those to whom it is offered. Consequently, treatment outcomes are based on the original allocation of patients, even if the clinician or owner ultimately decides not to follow the original experimental protocol. Results are typically subjected to an **intention-to-treat analysis** to assess outcomes.

Per-protocol and intention-to-treat analyses are discussed further in Section 8.2 above.

8.3.6 Assessment of Outcome

The perceptions and behavior of the participants (clinical investigators and clients) in a clinical trial may be affected systematically (biased) if they know who received which treatment. This is not a problem when the outcome is unequivocal, such as life or death. However, many clinical outcomes are subject to the interpretation of the observers. The rigor with which a patient is examined and the objectivity of the observers may be influenced by prior knowledge of an animal's treatment status. Clients may be anxious to see improvement in their pets or please the clinician. Clinicians may be more thorough in their examination of one group versus another. These are forms of what is sometimes referred to as **verification**, **detection,** or **workup bias**. A similar bias, **performance bias**, occurs when prior knowledge of which group animals belong to results in differences in care levels, making it difficult or impossible to conclude that a drug or other intervention caused an effect, as opposed to level of care. These sources of bias can be avoided by "**blinding**" the owners, the clinicians, or both to the treatment status of individual patients. Owners can be blinded by dispensing a placebo for control group patients. Clinicians can be blinded by use of a placebo or by not informing them of an animal's treatment status.

EXAMPLE 8.3: HOW EFFECTIVE IS ACUPUNCTURE FOR PAIN MANAGEMENT IN DOGS?

Background: Few high-quality veterinary medical studies have evaluated the effects of acupuncture (AP) in treating pain and improving quality of life in dogs.

Objectives: Silva et al. (2017) conducted a study to evaluate the efficacy of AP and related techniques alone or in combination with analgesics in chronic pain and quality of life of dogs with neurological and musculoskeletal diseases, using pre-validated scales answered by owners.

Study Design: Uncontrolled clinical trial.

Methods: Animals received one of two treatment combinations that were assessed by owners for up to 24 weeks in 181 dogs with neurological and musculoskeletal diseases

- Alternative medicine (ALG, $n = 50$), which included AP and related techniques (electroacupuncture, laserpuncture, ozone therapy, and/or, less frequently, pharmacopuncture [injection of microdoses of drugs into acupoints] or moxibustion [the stimulation of an acupoint by burning a cylinder of moxa placed close to the acupoint]), or
- Alternative medicine associated with conventional analgesics and adjuvant analgesics (AAG; $n = 131$). In this group, analgesics (nonsteroidal and steroidal anti-inflammatory drugs, opioids, amitriptyline, amantadine, gabapentin) and adjuvant analgesics (nutraceuticals, transcutaneous electrical nerve stimulation, magnetic and antalgic physical therapy) were used alone or in combination, and the protocols were discontinued or modified according to the individual clinical response.

Group assignments and treatment protocols were adjusted individually according to the specific needs of each patient and owner preferences. For ethical reasons, a placebo was not included as a negative control group. Treatment success was measured through weekly responses of owners to four questionnaires examining pain, locomotion, and health-related quality of life of their pets. The scores before and after the onset of treatment were evaluated using the Wilcoxon test and the evolution of success was evaluated by Kaplan–Meier curves. Differences were considered significant at $p < 0.05$. Some owners had difficulty interpreting survey questions, and in approximately 15% of cases the respondent differed between sessions.

Results: Although dogs with musculoskeletal diseases improved faster than those with neurological diseases according to some assessment scores, no statistically significant difference between treatments was found by Kaplan–Meier survival analysis when neurological and musculoskeletal diseases were grouped, or when each disease complex was analyzed separately.

Conclusions and Significance: Although no difference between treatments was found, the authors concluded that AP is an important conservative therapeutic tool to be included in the multimodal treatment protocols of neurological and musculoskeletal diseases in dogs.

FOLLOW-UP QUESTION 8.3

What source(s) of bias may have limited the ability of the authors to detect statistically significant differences between treatments in this study? (*Hint:* review items 8.3.1–8.3.8 for clues.) See Answer 8.3 at the end of this chapter.

8.3.7 Placebo Effect

A **placebo** is defined as any medical intervention that has a nonspecific, psychological, or psychophysiologic therapeutic effect, or that is used for a presumed specific therapeutic effect on a patient, symptom, or illness but is without specific activity for the condition being treated (McMillan, 1999). It follows that the **placebo effect** is the nonspecific psychological or psychophysiologic therapeutic effect induced by a placebo. The effect may be positive or negative, e.g., favorable or unfavorable. Placebos are important both as a control in clinical trials and for understanding the mechanism of how they work.

A possible mechanism of action of a placebo in animal subjects is through the effect of human contact (visual and tactile) on animal health. Among human observers, expectations of a response may influence the subjective interpretation of the results of animal studies and erroneously attribute a response to either the placebo or treatment. This is really a form of **investigator bias** rather than a biologically mediated effect.

In clinical trials in which a placebo is selected as the control method, it may be useful to include a second control group in which a placebo is not administered. This would permit placebo effects to be distinguished from other causes of disease resolution.

8.3.8 Statistical Analysis

Many reports of clinical trials end by concluding that a treatment offered a "significant" improvement over existing techniques or controls. Whenever this word is used, it should be backed up by appropriate statistical analysis, and it should be stated at the outset how the results were analyzed. Statistical tests must answer one fundamental question: *How certain can we be that the observed results did not arise by chance alone?*

Statistical significance does not automatically equate with **clinical significance**. As the number of animals in each comparison group increases, the statistical significance of differences in group means or medians also tends to increase. However, if there is considerable overlap among individuals across comparison groups, then we may not be able to accurately predict clinical outcomes for individual patients.

Statistical significance does not automatically equate with clinical significance.

EXAMPLE 8.4: HOW EFFECTIVE IS PROACTIVE ANTI-INFLAMMATORY THERAPY FOR THE LONG-TERM MANAGEMENT OF CANINE ALLERGIC DERMATITIS?

Background: Canine atopic dermatitis (CAD) is a common, highly pruritic disease with an unpredictable course and frequent flares of inflammation. Long-term remission between flares can be difficult to achieve. Therefore, additional strategic forms of treatment, such as proactive therapy, are needed in order to target flare prevention. Proactive therapy begins with intensive topical anti-inflammatory therapy until lesions are in remission, followed by long-term, low-dose, intermittent application of the anti-inflammatory agent to the previously affected skin.

Objectives: Lourenço et al. (2016) evaluated the efficacy of a long-term, proactive, intermittent treatment regimen with a 0.0584% hydrocortisone aceponate (HCA) spray among 41 client-owned dogs with spontaneous atopic dermatitis (AD).

Study Design: Randomized controlled clinical trial.

Methods: The study was conducted as a randomized, placebo-controlled, double-blinded clinical trial with an end-point of treatment failure. The clinical diagnosis (case-definition) of AD was made according to accepted criteria and after ruling out other causes of pruritus. Dogs were treated once daily with HCA spray to remission, then randomly assigned to receive either the HCA spray ($n = 21$) or a placebo ($n = 20$) spray on two consecutive days each week. Group assignment was masked from the owners and investigator until the trial was completed. The HCA spray and the placebo were supplied in identical pre-packaged bottles labeled A, B, C, or D. A student or a nurse dispensed the bottles to the owners and maintained the corresponding records. A single investigator who was not involved in the treatment allocation assessed all of the treatment outcomes. All dogs were on appropriate flea control. No topical or systemic anti-inflammatory or antimicrobial agents were permitted. Intention-to-treat (ITT) analysis was used to evaluate outcomes. At Day 0, all dogs were in remission or had mild AD based on their Canine Atopic Dermatitis Extent and Severity Index, version 3 (CADESI-03) scores. Custom-made journal forms were given to the owners to record treatment applications, unexpected occurrences, or adverse effects. Regular telephone calls to the owners were made to obtain updates on the treatment plan and the dog's condition. The study was concluded after 12 months, at which time the owners were asked to return all bottles (empty or not) so that compliance could be assessed. Four dogs were lost to follow-up and four were withdrawn after receiving prohibited medication. The Kaplan–Meier method was used to estimate the distribution of time to relapse of AD.

Results: The time to relapse (Figure 8.2) was significantly longer in the HCA group (median 115 d; range 31–260 d) compared with the placebo group (median 33 d; range 15–61 d) ($p < 0.0001$). No adverse events were attributable to the HCA spray.

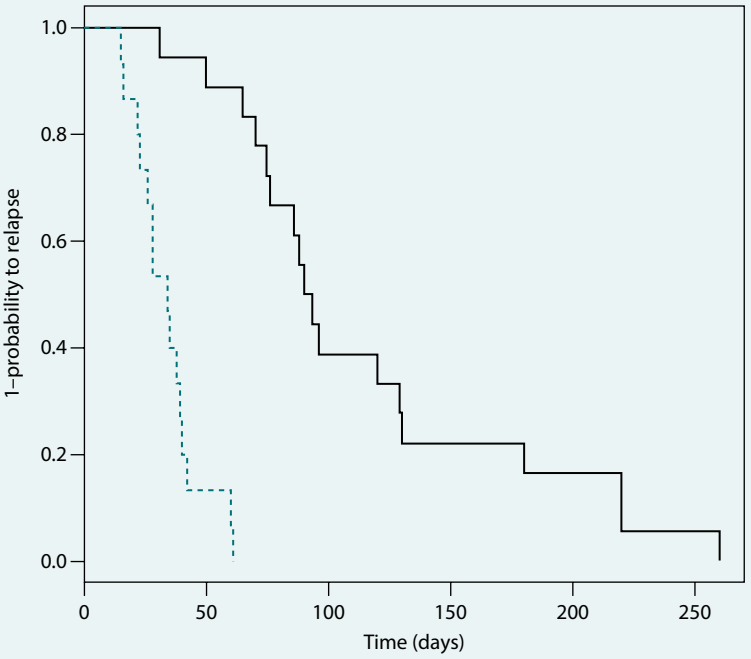

FIGURE 8.2 Kaplan–Meier survival analysis comparing times to relapse in the therapy of canine atopic dermatitis: ---, placebo group; —, hydrocortisone aceponate spray (HCA) group. The relapse-free interval was significantly longer in the HCA group (median 115 d; range 31–260 d) compared with the placebo group (median 33 d; range 15–61 d) ($p < 0.0001$). (From Lourenço AM et al. *Vet Dermatol* 2016;27:88-e25. With permission.)

Conclusions and Significance: The authors concluded that proactive long-term therapy of CAD with an HCA spray administered on two consecutive days each week is effective and well tolerated.

FOLLOW-UP QUESTION 8.4

This study was especially well designed and controlled. What sources of bias described in 8.3.1–8.3.8 were addressed in this clinical trial? See Answer 8.4 at the end of this chapter. (*Hint:* see Figure 8.1 and Table 8.1 for clues.)

8.4 SUBGROUPS

During the analysis of a clinical trial, the investigators may be tempted to compare outcomes among specific subgroups of patients. If the number of patients in the clinical trial is large, then the number of individuals in each subgroup may be adequate for meaningful comparisons, provided that systematic differences among the groups being compared are adjusted for. However, as the number of subgroup comparisons increases, so does the likelihood that a statistically significant difference will be detected, even if it is not real. Validity of findings from subgroups is not a problem unique to clinical trials. Clinical studies of frequency, risk, prognosis, and cause often include the frequency of findings in various subgroups. For example, in their comparison of ureteral stents versus ureterotomy for the relief of ureteral blockage in cats (Example 8.2 above), the authors reported that cats that developed abdominal effusion after surgery (6 cats in the ureteral stenting group, 12 cats in the ureterotomy group) were significantly ($p = 0.003$) less likely to be discharged from the hospital.

The number of cats involved in the study permitted the detection of a statistically and clinically significant finding among subgroups.

> ***As the number of subgroup comparisons increases, so does the likelihood that a statistically significant difference will be detected, even if it is not real.***

8.5 CLINICAL TRIALS IN PRACTICE

Randomized controlled clinical trials are the best available means for assessing the value of treatment. However, because of many practical limitations, the majority of therapeutic questions are answered by other means, particularly uncontrolled and nonrandomized trials. The need to administer some sort of treatment is largely responsible for the large percentage of case reports and uncontrolled clinical trials.

REFERENCES

Berruex F, Gerber V, Wohlfender FD et al. Clinical course of sarcoids in 61 Franches-Montagnes horses over a 5–7 year period. *Vet Q* 2016;36:189–196. Full text available at: https://www.tandfonline.com/doi/full/10.1080/01652176.2016.1204483

Culp WTN, Palm CA, Hsueh C et al. Outcome in cats with benign ureteral obstructions treated by means of ureteral stenting versus ureterotomy. *J Am Vet Med Assoc* 2016;249:1292–1300.

Fletcher RH, Fletcher SW, and Fletcher GS. *Clinical Epidemiology—The Essentials*, 5th ed. Baltimore: Lippincott Williams & Wilkins, Baltimore; 2014.

Fletcher RH, Fletcher SW, and Wagner EH. *Clinical Epidemiology—The Essentials*, 1st ed. Baltimore: Williams and Wilkins; 1982.

Gordis L. *Epidemiology, 5th ed*. Philadelphia: Elsevier-Saunders; 2014.

Lourenço AM, Schmidt V, São Braz B et al. Efficacy of proactive long-term maintenance therapy of canine atopic dermatitis with 0.0584% hydrocortisone aceponate spray: A double-blind placebo controlled pilot study. *Vet Dermatol* 2016;27:88–e25. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/vde.12285

McMillan FD. The placebo effect in animals. *J Am Vet Med Assoc* 1999;215:992–999.

Olivry T and Mueller RS. Evidence-based veterinary dermatology: A systematic review of the pharmacotherapy of canine atopic dermatitis. *Vet Dermatol* 2003;14:121–146.

Sackett DL. Bias in analytic research. *J Chronic Dis* 1979;32:51–63.

Silva NEOF, Luna SPL, Joaquim JGF et al. Effect of acupuncture on pain and quality of life in canine neurological and musculoskeletal diseases. *Can Vet J* 2017;58:941–951. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5556488/

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 8.1: Results of this study should be considered when critically assessing treatment effects of therapies directed against equine sarcoid, especially uncontrolled studies. The clinical course of ES is such that more than half of affected horses may undergo an uneventful regression of lesions regardless of treatment. Consequently, uncontrolled clinical trials of any therapeutic regimen for the disease are likely to be favorable. This example also illustrates the hazards of using "testimonials" or case reports to guide treatment decisions.

Answer 8.2: This is an example of "**intention-to-treat analysis**." See Section 8.2 above for further discussion of the rationale and benefits of this strategy.

Answer 8.3: According to the authors, "a limitation inherent to any clinical experiment is the heterogeneity of the population and epidemiological data, with different severity of diseases and

associations among them." Consequently, some variables could not be controlled. Perhaps the greatest challenge to interpreting the results of this study is that, according to the authors, there were no defined criteria for assigning animals to treatment groups, making it impossible to compare treatment efficacy between the different treatments. Random assignment to treatment groups and consistent treatment protocols could have improved the validity of the results. Furthermore, owners were not "blinded" to the treatment status of their pets, which might bias their responses to the subjective assessment of treatment outcomes. The use of placebo controls would have addressed this issue, although this option is often difficult to implement in a practice environment. Another option would be to use historical cohorts treated by other means as controls.

Answer 8.4: Look for these key words in the Example 8.4 abstract: **case-definition, placebo-, double-blinded, randomized, intention-to-treat (ITT) analysis, *p*-values**. Selection bias in breed, age, sex, weight, and clinical severity was not apparent. Randomized treatment allocation was made according to a predetermined allocation code. Detection (workup) bias by the investigators was unlikely, as they were blinded to treatment allocation, and a dispenser who did not participate in any outcome assessments performed treatment-related follow-up. Performance bias (systematic difference in care between groups) was considered unlikely, as concomitant treatments were predefined, stabilized before the trial, maintained during the trial, and were similar between the placebo and HCA groups. Attrition bias (a systematic error caused by unequal loss of participants from a randomized controlled trial) was potentially present, with eight dogs withdrawn from the two phases of the study; however, ITT analysis reduced the likelihood that this kind of bias would influence the conclusions drawn from the study. The highly significant p-value ($p < 0.0001$) made it unlikely that the observed difference between treatment groups was due to chance.

9 Statistical Significance

9.1 INTRODUCTION

"Figures don't lie but liars can figure." —Anonymous

"There are three types of lies: lies, damn lies and statistics." —Mark Twain

"Torture numbers and they'll confess to anything." —Gregg Easterbrook in *The New Republic*

Statistical analyses, once a rarity in medical journals, are now routinely encountered in the medical literature, and veterinary journals are no exception. Statistical analyses often have immense practical importance since research results are frequently the basis for decisions about patient care. If the choice of treatment hinges on faulty statistics, a great deal of harm may be done. An effective treatment may be dismissed as worthless and an ineffective treatment may be adopted. By learning to recognize statistical errors in the veterinary literature, practitioners can protect themselves, their practices, and their patients from the harm that may result when invalid study results are accepted and applied.

Besides treatment outcomes, statistics are used to confirm or refute the significance of risk and prognostic factors, and as a quality-control component in population surveys. The likelihood of failing to detect disease in a population depends not only on the properties of diagnostic tests being used, but also on the degree to which the sample represents the population as a whole. Thus, all aspects of the practice of medicine require that statistics be used, and that they be used correctly.

Until now we have used **descriptive statistics** (measures of central tendency and dispersion) to describe clinical data. We now turn to **inferential statistics** to help us determine whether observed outcomes are real or the result of random variation.

Statistical analyses are now much easier to perform than in the past. Many basic statistical functions are built into smartphone apps, while others are available as personal computer spreadsheet programs and specialized software packages. Statistical errors are not uncommon in medical research. Since most investigators rely on preprogrammed statistical packages, the most frequent statistical errors arise from analyses that are inappropriate for the type of data or study design, rather than "errors of execution." In this chapter, we discuss the application and interpretation of statistical tests in clinical epidemiology and the rules that guide the selection of appropriate statistical tests.

Statistical analyses, once a rarity in medical journals, are now routinely encountered in the medical literature, and veterinary journals are no exception.

9.2 HYPOTHESIS DEFINITION AND TESTING: AN OVERVIEW

In this chapter, many of the details of the design and analysis of scientific research are discussed from the perspective of statistical testing. The primary purpose of statistical testing is to determine whether the observed results are real or could have occurred by chance. Before embarking on the

details, it may be useful to provide a brief overview of hypothesis testing and introduce some of the major concepts. Each will be discussed in greater detail in the pages that follow.

9.2.1 The Steps in Hypothesis Testing: An Example

Any scientific investigation, epidemiologic or otherwise, begins with a research question, e.g., the objective or purpose of the study. The initial research question may reflect a general concern to be restated as one or more specific research questions. For example, an initial research question might be: ***Is the widespread use of antimicrobials by humans responsible for the presence of antimicrobial resistant bacteria (ARB) in wildlife (Swift et al., 2019)*?** More specific questions might ask whether particular kinds of exposures: sewage treatment plants (***STPs***), farm sites (***Farm***), and sites with no sources of waste containing anthropogenic ARB or antimicrobials (***Central***) are associated with the differences in the prevalence of ARB among wildlife in these areas.

The next step is to formulate a **research hypothesis** that summarizes the elements of the study: the sample, the design, and the predictor and outcome variables. The research hypothesis should establish the basis for tests of statistical significance. This is usually done by restating the research hypothesis in the form of null and alternative hypotheses. The **null hypothesis** states that there is no association between the predictor and outcome variables. In the ARB in wildlife example, this might be stated as: *when comparing the effect of site (Farm, Central, and STP), season (summer and autumn) and taxa (bird or mammal), the null hypothesis* (H_0) *states that there is no difference in the prevalence of ARB bacteria among the populations, and the alternative hypothesis* (H_1) *states there are differences among the populations.* **The alternative hypothesis cannot be tested directly; it is accepted by default if the test of statistical significance rejects the null hypothesis** (see below).

Research hypotheses are usually stated as either directional or non-directional. A **directional (one-sided) hypothesis** of the ARB in wildlife example would state that the prevalence of ARB in wildlife living in *STP* or *Farm* areas is greater than among wildlife living in *Central* areas. A **non-directional (two-sided) hypothesis** would simply state that there is an association between exposure and outcome without specifying whether exposed wildlife are more or less likely to harbor ARB. The practical significance of choosing between a directional and non-directional hypothesis lies in the fact that the non-directional hypothesis is more stringent; i.e., the evidence (data) required to reject the null hypothesis must be stronger for a non-directional hypothesis than with a directional hypothesis. Non-directional hypotheses also require a larger sample size. For these reasons, non-directional hypotheses are generally preferred when estimating the required sample size and analyzing the data.

Once the data are analyzed, statistical tests determine the ***p*-value**, the probability or likelihood of obtaining the observed or more extreme results by chance alone if the null hypothesis were true. *p*-values are expressed as **one-tailed** or **two-tailed** in accordance with whether the hypotheses being tested are directional or non-directional, respectively. The null hypothesis is rejected in favor of the alternative hypothesis if the *p*-value is less than the predetermined level of statistical significance. By convention this is usually 5%, i.e., we are willing to erroneously conclude that an association between predictor and outcome variables exists up to 5% of the time. Statistical tests thus give us an idea of the level of confidence that we can have in our results.

9.2.2 Results and Conclusions

Returning to the original research question above, i.e., the role of anthropogenic factors in the patterns of antimicrobial resistance (AMR) found in wildlife, Swift et al. (2019) reported that the overall prevalence of ARB (*Eschericha coli*) among wildlife was 54% ($n = 262$) and was significantly explained by a binomial logistic regression model that included season, taxa, and site. ARB prevalence in samples from the STP site was 61.3%, which was significantly higher ($p = 0.029$) than the prevalence of resistance in samples from the Central site (50.0%). Prevalence in samples from the Farm site was 52.1% and did not significantly differ ($p = 0.28$) from that in

Central site samples. *Escherichia coli* from samples collected in summer (prevalence = 65.4%) were significantly more likely ($p < 0.0001$) to be resistant than those collected in autumn (36.9%). There was a "tendency" ($p = 0.056$) for mammalian fecal samples to have a higher prevalence (55.7%) of resistant *E. coli* than avian samples (40.7%). The authors concluded that antimicrobial resistance in commensal bacteria of wildlife is not driven simply by anthropogenic factors and that this may limit the utility of wildlife as sentinels of spatial variation in the transmission of environmental AMR.

9.3 INTERPRETATION OF STATISTICAL ANALYSES

Many of the rules that apply to the interpretation of statistical tests are similar to those discussed earlier in the context of diagnostic tests. In the usual situation, the outcome of clinical studies is expressed in dichotomous terms: *either a difference exists, or it doesn't*. Since we are using samples to predict the true state of affairs in the population, there always exists a chance that we will come to the wrong conclusion. When statistical tests are applied, there are four possible conclusions—two are correct and two are incorrect (Figure 9.1).

Two of the four possibilities lead to correct conclusions—either a real difference exists (cell a) or it does not (cell d). There are also two ways of being wrong. **Alpha or Type I error** (cell b) results when we conclude that outcomes are different when, in fact, they are not. Alpha error is analogous to the false-positive result of diagnostic tests. **Beta or Type II error** (cell c) occurs when we conclude that outcomes are not different when, in fact, they are. Beta error is analogous to the false-negative result of diagnostic tests.

When statistical tests are applied there are four possible conclusions—two are correct and two are incorrect.

9.3.1 Concluding a Difference Exists

9.3.1.1 The Null Hypothesis

Statistical tests reported in the medical literature are usually used to disprove the null hypothesis that no difference exists between groups. If differences are detected, they are reported with the corresponding *p*-value, which expresses the probability of obtaining the observed (or more extreme) result under the assumption that the null hypothesis is true, e.g., by chance. This *p*-value is sometimes referred to as "p_a" to distinguish it from beta error.

Conclusion of statistical test	True difference: Present	True difference: Absent
Different (reject null hypothesis)	(a) Correct	(b) Incorrect (Type I or alpha error)
Not different (accept null hypothesis)	(c) Incorrect (Type II or beta error)	(d) Correct

FIGURE 9.1 The relationship between the statistical analysis of study results and the true difference between possible outcomes.

9.3.1.2 Statistical Significance

A p-value is usually considered statistically significant if it falls below 0.05; i.e., we are willing to be wrong up to 5% of the time. Since not everyone agrees with this criterion, it is preferable to specify the actual probability of an alpha error, such as $p = 0.10$, $p = 0.005$, etc.

Earlier in the book (Example 7.1), a distinction was made between statistically and clinically significant findings based on a report by Trefz et al. (2017). Despite the fact that intravenous administration of a hypertonic sodium bicarbonate solution induced a statistically significant decline ($p = 0.003$) in plasma potassium concentration compared with a hypertonic sodium chloride solution in hyperkalemic diarrheic calves, clinical findings such as posture, behavior, and strength of the suckling reflex were not significantly different among treatment groups. In other words, blood biochemistries were not reflected in clinical status of patients.

A similar discrepancy between p-values and clinical significance can result when there is significant overlap of data points among groups being compared. The p-value does not convey the magnitude of the difference between groups, only the likelihood that a difference of that magnitude could have arisen by chance alone. If individual animal variability is such that considerable overlap occurs between groups, the difference in group means could be statistically significant but not clinically relevant.

The p-value does not indicate the magnitude of the difference between groups, only the likelihood that a difference of that magnitude could have arisen by chance alone.

EXAMPLE 9.1: HOW USEFUL ARE ACUTE PHASE PROTEINS (APPS) FOR DISTINGUISHING FELINE INFECTIOUS PERITONITIS (FIP) FROM OTHER DISEASES?

Background: Feline infectious peritonitis is a lethal infectious disease that can occur in two clinically distinct forms, the more common effusive (wet) form and the granulomatous (dry) form. Ascites or pleural effusion due to FIP have to be differentiated from other potential causes such as cardiac disease, neoplasia, or septic effusion. Although several diagnostic tests have been developed to diagnose FIP, differentiation between FIP and diseases with similar clinical presentation remains challenging in clinical settings.

Objectives: Example 4.1 presented the results of a study by Saverio et al. (2007) on the diagnostic utility of serum α1-acid glycoprotein for feline infectious peritonitis. In a follow-up study, Hazuchova et al. (2017) compared the clinical utility of AGP with two other acute phase proteins, serum amyloid A (SAA) and haptoglobin (Hp) as a diagnostic tool to differentiate between feline infectious peritonitis and other diseases in cats with body cavity effusions.

Study Design: Cross-sectional.

Methods: Cats with pleural, abdominal, or pericardial effusion were prospectively enrolled in the study and classified as having or not having FIP. Cats without FIP were further subdivided into three subgroups: cardiac disease, neoplasia, and other diseases. Serum amyloid A, haptoglobin, and α1-acid glycoprotein were measured in serum and effusion using assays previously validated in cats. Serum and effusion samples were available for the measurement of APPs from 88 and 67 cats, respectively. Data were found not to be normally distributed, so the numerical values were expressed as median and range and non-parametric statistical tests (see below) used to assess the statistical significance of differences between groups. For each parameter tested, a receiver operating characteristic curve and the area under the curve were calculated.

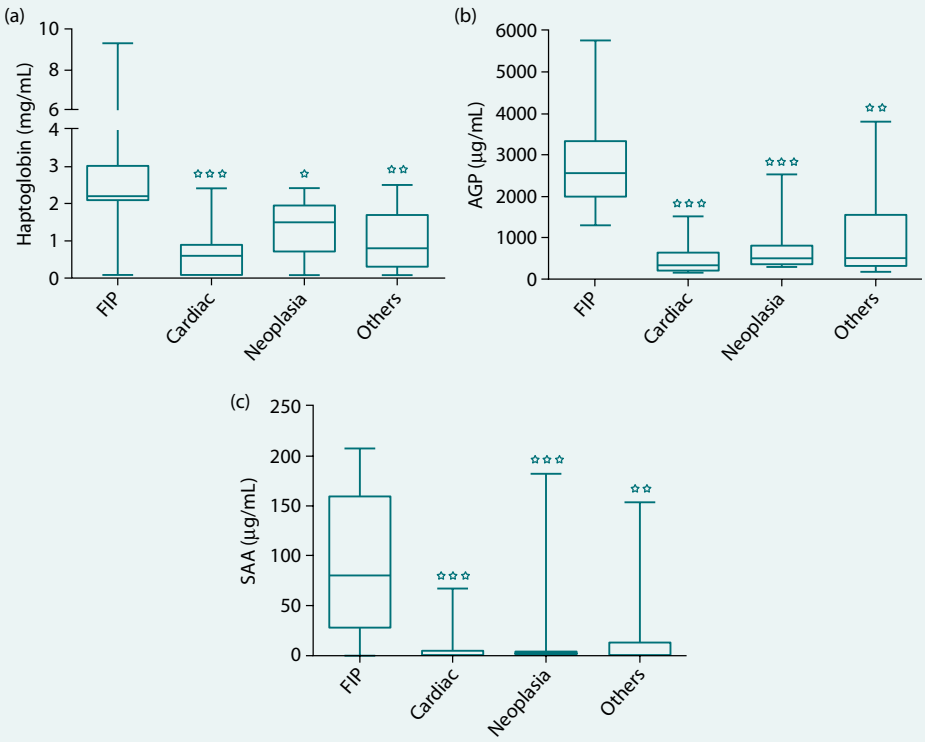

FIGURE 9.2 (a) Haptoglobin, (b) α1-acid glycoprotein (AGP), and (c) serum amyloid A (SAA) concentration in effusion of cats with feline infectious peritonitis (FIP; n = 14), cardiac disease (n = 17), neoplasia (n = 21), and other diseases (n = 15). The boxes represent the 25th and 75th quartiles, with a horizontal line at the median. The whiskers represent the range of the data. Stars represent the significance levels (***$p < 0.001$, **$p < 0.01$, *$p < 0.05$) when comparing the group with cardiac disease, neoplasia, and other diseases with the FIP group. Although differences between FIP and non-FIP groups were statistically significant, they were not clinically significant due to the degree of overlap (range) of values for individual cats. (From Hazuchova K et al. *J Feline Med Surg* 2017;19:809–816. With permission.)

Results: Concentrations of the APPs in serum and effusion were significantly different in cats with and without FIP ($p < 0.001$) for all three APPs. However, there was considerable overlap of individual data points (Figure 9.2). The best APP to distinguish between cats with and without FIP was AGP in the effusion; a cut-off value of 1550 µg/mL had a sensitivity and specificity of 93% each for diagnosing FIP.

Conclusions and Significance: AGP, particularly if measured in effusion, was found to be useful in differentiating between FIP and other diseases, while SAA and Hp were not. In some diseases (e.g., septic processes, disseminated neoplasia) the concentration of all three APPs was as high as in cats with FIP (Figure 9.2). Therefore, none of these can be recommended as a single diagnostic test for FIP.

FOLLOW-UP QUESTION 9.1

The findings of this study suggest that none of the three APPs would be clinically useful in the differential diagnosis of FIP. How could their diagnostic utility be improved? See Answer 9.1 at the end of this chapter.

9.3.1.3 Confidence Limits

Confidence limits are the numbers at the upper and lower end of a confidence interval; for example, if your mean is 7.4 with confidence limits of 5.4 and 9.4, your confidence interval is 5.4 to 9.4 (McDonald, 2014). The **confidence interval (CI)** provides a way of expressing the range over which a value is likely to occur. This value could be the difference between the means of two groups, or the theoretical range over which a measurement, such as blood pressure, might occur. Although any range can be used, the 95% confidence interval is most commonly used in the medical literature. It means that the probability of including the true value within the specified range is 0.95.

Confidence intervals have appeared in a number of examples throughout the book. See, for example, estimates of normal ranges for canine serum creatinine values (Table 2.4), ROC analysis of α1-acid glycoprotein for the diagnosis of feline infectious peritonitis (Figure 4.2a and b), the incidence density rate of *Anaplasma marginale* infection among groups of cattle (Example 5.5), odds ratios for contracting psittacosis from wild and domestic birds (Example 6.6), and estimates of canine longevity (Example 7.3 and Figure 7.3).

9.3.1.4 Confidence Interval for a Rate or Proportion

Confidence intervals used in descriptive statistics often derive the mean, variance, and standard deviation from measured interval-level values. Frequency measures such as incidence and prevalence present a special problem in that they are derived from dichotomous (nominal) outcomes (as presence or absence of disease) rather than measured values. One approach to estimating the confidence interval for such proportions is known as the **normal approximation** (McDonald, 2014). It considers the disease frequency value to be the mean and assumes that the sample proportions are normally distributed, yielding symmetrical confidence limits. According to this statistical model, the **variance** of disease prevalence = $[p\,(1-p)/\mathrm{n}]$, where p = proportion of affected individuals and n = sample size. The **standard deviation** of disease prevalence is equal to the square root of the variance. Since we are really estimating the standard deviation of the sampling distribution of a proportion (or mean) rather than the standard deviation of individual values around the mean, the derived value is called the **standard error of the proportion**.

EXAMPLE 9.2: WHAT IS THE BEST METHOD FOR ESTIMATING THE CONFIDENCE LIMITS FOR A RATE OR PROPORTION, PARTICULARLY FOR DISEASES OF LOW PREVALENCE?

For proportions near zero or one, the normal approximation to confidence intervals (see above) yields incorrect results. To illustrate this point, let's see what would happen if the above approach were used to estimate the confidence interval for a disease of low prevalence. The data is drawn from a study by Wolfe et al. (2018) evaluating the effectiveness of a test and cull strategy for reducing chronic wasting disease prevalence over a 6-year period in a naturally infected, free-ranging mule deer (*Odocoileus hemionus*) herd. At the initiation of the study, 3 of 66 sampled deer tested positive for CWD based on tonsil biopsy immunohistochemistry, yielding a prevalence of infection of 4.55%. If we were to use the normal approximation to estimate the 95% confidence interval for CWD prevalence from this data:

- The variance of CWD prevalence = $(0.0455 \times 0.9955) \div 66 = 0.000657$
- The standard error of the proportion (square root of the variance) = $\sqrt{0.000657}$ or $\approx 2.56\%$.
- The 95% confidence interval for CWD would be $4.55\% \pm (1.96 \times 2.56\%)$, or **−0.48% to 9.58%**

The fact that the above formula estimates that CWD prevalence could be less than 0%, even though infection was confirmed through tonsil biopsy, results from the fact that the normal approximation assumes a normal distribution around the mean.

Fortunately, there is a more accurate (but more complicated) formula based on the **binomial distribution** for calculating confidence limits for proportions (McDonald, 2014). The binomial distribution of proportions is not symmetrical around the mean, except for the special case where $p = 0.50$.

Lookup tables are available that give exact binomial confidence limits (bCL) for a proportion (Zwillinger and Kokoska, 2000). However, lookup tables lack precision, as the row and column intervals are not uniform for higher numbers and therefore do not cover the full range of possible sample data. An easy-to-use online calculator for more precisely estimating binomial confidence limits of a proportion can be found on GraphPad's QuickCalcs website at https://www.graphpad.com/quickcalcs/. Entering the above numerator (3) and denominator (66) values for CWD into the online calculator ("categorical data -> confidence interval of a proportion") yields the following results for the "exact" binomial confidence intervals:

- The 90% confidence interval extends from 0.0125 to 0.1133.
- The 95% confidence interval extends from 0.0095 to 0.1271.
- The 99% confidence interval extends from 0.0052 to 0.1566.

Note that the lower limit for the 95% confidence limit is now greater than 0. In fact, even when the numerator is 0, the lower confidence limit is never less than 0. Further, since the 95% confidence interval is not symmetrical around the mean (skewed to the right), you can't report the prevalence of CWD infected deer as "4.5% ± something." Instead, you'd have to say "4.55% with 95% confidence limits of 1% and 13.3%." The prevalence remains the same, but the 95% confidence limits are now 0.0095 and 0.1271, as reported by Wolfe et al. (2018) in their study.

FOLLOW-UP QUESTION 9.2

The confidence limits for prevalence of CWD estimated above are all rather broad, possibly limiting the ability to detect small changes in CWD prevalence resulting from a test and cull strategy. What can be done to improve the chances of detecting a statistically significant change in prevalence? See Answer 9.2 at the end of this chapter.

9.3.1.5 One-Tailed versus Two-Tailed Tests of Significance

When performing a statistical test, we may be given the option of choosing a one- or two-tailed test of significance. The *p*-values will differ depending on which is chosen. If we are certain that differences can only occur in one direction, then a one-tailed test can be used. Examples might be whether an observed temperature rise or drop in erythrocyte count deviated significantly from normal. If a difference could occur in either direction, then a two-tailed test should be used. Two-tailed tests are more conservative, i.e., the difference required for statistical significance must be greater than with one-tailed tests. On the other hand, one-tailed tests are more likely to detect true differences when they occur. See Figures 2.9 and 2.10 for a comparison of one- and two-tailed cutoffs.

9.3.2 Concluding a Difference Does Not Exist

9.3.2.1 Statistical Significance

By default, *p*-values ≥ 0.05 imply that no difference between outcome or treatment groups exist. Actually, a $p \geq 0.05$ does not mean that one factor is comparable to, equivalent to, or not different

from the second factor. All that has been demonstrated is an absence of evidence of a difference (Christley and Reid, 2003). In other words, failing to reject the null hypothesis does not mean that we have proven it. There is a chance that a true difference occurred but we failed to detect it because of poor study design, inadequate numbers of individuals, or bad luck. The probability of this kind of error, known as **beta or Type II error**, is expressed as $\mathbf{P_b}$.

9.3.2.2 Power

Power is the probability that a study will find a statistically significant difference when one exists. Power is analogous to diagnostic test sensitivity and is related to beta error by the equation

$$\text{Power} = 1 - P_b$$

P_b is the major determinant of sample size in epidemiologic research. Whereas alpha error is generally set at <5%, beta error is generally set at 20%. Thus, when viewed as a diagnostic test, statistical criteria for determining sample size are more specific than sensitive. The determination of sample size is further discussed later in this chapter.

9.3.3 Concluding an Association Exists

9.3.3.1 Agreement between Tests

As defined in Chapter 3, **concordance** is the proportion of all test results on which two or more different tests or observers agree. The level of agreement is frequently expressed as the **kappa (*k*) statistic** (also referred to as **"Cohen's *k*"**), defined as the proportion of potential agreement beyond chance exhibited by two or more tests. Expected agreement by chance is calculated by the method of **marginal cross-products** (see Table 9.1 and Example 9.3). The value of kappa

TABLE 9.1

Concordance between Macroscopic Diagnosis (Rows) and Microscopic Report (Columns) for Dogs Undergoing Gastric Endoscopy

		Histology		
		Acute gastritis	**Chronic gastritis**	**Total**
Gastroscopy	Acute gastritis	(a) 5	(b) 15	20
	Chronic gastritis	(c) 6	(d) 76	82
	Total	11	91	102

Steps in the calculation of the kappa (*k*) statistic from the above data.

$$\text{Observed agreement(concordance)} = \frac{(\text{Observed a}) + (\text{Observed d})}{a+b+c+d} = \frac{(5+76)}{102} = 79.41\%$$

$$\text{Expected (chance) agreement for cell a} = \frac{(a+b)\times(a+c)}{a+b+c+d} = \frac{(20\times11)}{102} = 2.16$$

$$\text{Expected (chance) agreement for cell d} = \frac{(c+d)\times(b+d)}{a+b+c+d} = \frac{(82\times91)}{102} = 73.16$$

$$\text{Expected (chance) agreement overall} = \frac{(\text{Expected a}) + (\text{Expected d})}{a+b+c+d} = \frac{(2.16+73.16)}{102} = 73.84\%$$

$$\text{Agreement beyond chance (kappa)} = \frac{\text{Observed agreement} - \text{Expected agreement}}{100\% - \text{Expected agreement}} = \frac{(79.41\% - 73.84\%)}{(100\% - 73.84\%)} = 0.213$$

Source: Marchesi MC et al. *Veterinaria Italiana* 2017;53:309–313.

Note: The 95% confidence interval for kappa: –0.016 to 0.442.

ranges from −1.0 (perfect disagreement) through 0.0 (chance agreement only) to +1.0 (perfect agreement). Although the interpretation of kappa results varies among statisticians, the following scale has proven useful: <0.2 poor agreement, 0.21–0.4 fair, 0.41–0.6 moderate, 0.61–0.8 strong, and more than 0.8 near complete agreement (McHugh, 2012). It should be pointed out that percent concordance and the kappa statistic do not tell us which test is correct, only the level of agreement between them.

The following example illustrates the clinical utility of the kappa statistic for expressing the level of agreement between methods or observers.

EXAMPLE 9.3: WHAT IS THE LEVEL OF AGREEMENT BETWEEN ENDOSCOPIC AND HISTOLOGICAL METHODS FOR THE DIAGNOSIS OF GASTRIC DISEASE IN THE DOG?

Background: Gastric diseases are common in dogs. Endoscopy is a common, minimally invasive diagnostic technique used to observe internal organs, e.g., the stomach, and to obtain mucosal biopsy samples for histopathological examination.

Objectives: Marchesi et al. (2017) evaluated the degree of concordance between endoscopic and histological evaluation of gastric diseases in dogs.

Study Design: Cross-sectional.

Methods: Medical records from 129 dogs that had undergone gastroscopy at the Veterinary Hospital of Perugia University (Perugia, Italy) because of vomiting, lack of appetite, and weight loss between 2009 and 2012 were reviewed. Gastroscopy was performed on all patients after general anesthesia. Findings were first classified according to the macroscopic view as acute gastritis, chronic gastritis, or nodular gastropathy. Three biopsies of each gastric region were obtained. The same pathologists reviewed all slides, and cases were classified according to histological presentation as acute gastritis, chronic gastritis, or gastric tumors. The agreement between endoscopic and histological reports of acute and chronic gastritis or gastric tumors was assessed by Cohen's *k* coefficient. Considering histological diagnosis the "gold standard," the authors also calculated sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the endoscopic report.

A subset of the authors' data appears in Table 9.1 and is used to illustrate how concordance and the kappa statistic are calculated.

Results: Endoscopy showed a sensitivity of 45%, 88%, and 100% for acute gastritis, chronic gastritis, and gastric tumors, respectively, and specificity of 84%, 71%, and 100%. The PPV and NPV were 25% and 93% for acute gastritis, 93% and 60% for chronic gastritis, and 100% and 100% for gastric tumors, respectively. Test concordance was 79.41%. On the basis of column and row totals, we would expect the two tests to agree 73.84% of the time by chance alone, and the remaining potential agreement beyond chance would therefore be 26.16% (100% – 73.84%). The observed agreement beyond chance was 5.57% (79.41% – 73.84%), yielding a value for kappa of 0.213 (5.57% ÷ 26.16%). Based on the criteria outlined above, a kappa value of 0.213 reflects a 'fair" level of agreement (concordance) between the diagnostic procedures.

Conclusions and Significance: When all cases of gastric disease were analyzed by the authors, a value for kappa of 0.35 (95% CI: 0.14–0.56) was obtained. This was an improvement over the subset of data analyzed above, but still reflects only a "fair" level of agreement between diagnostic methods. The authors concluded that gastric endoscopy cannot be relied upon as a screening test, and that both endoscopic and histological exams should be conducted to optimize diagnosis.

FOLLOW-UP QUESTION 9.3

The authors considered histological diagnosis the "gold standard" and used it to calculate sensitivity, specificity, positive predictive value, and negative predictive value of the endoscopic results. Which of these results is subject to the prevalence of disease in the study group? (*Hint:* see Chapter 3, "Evaluation of Diagnostic Tests for a clue.") See Answer 9.3 at the end of this chapter.

9.3.3.2 Association between Two Variables

Statistics are also used to describe the degree of association between variables. The **Pearson product-moment correlation coefficient**, or **Pearson *r***, is a measure of the strength and direction of a linear (straight line) association between two interval-level variables. Examples might be the relation between body weight and blood volume, or between biochemical or physiologic measures such as blood packed cell volume and hemoglobin concentration. The value of r may take any value between −1 and 1. If r is either −1 or 1, the variables have a perfect linear relationship. If r is near −1 or 1, there is a high degree of linear correlation. A positive correlation means that as one variable increases, the other also increases. A negative correlation means that as one variable increases, the other decreases. If r is equal to 0, we say the variables are uncorrelated and that there is no linear association between them.

The correlation coefficient is the square root of the **coefficient of determination**, r^2, which expresses the amount of variation in the data that is accounted for by the linear relationship between two variables and may take any value between 0 and 1. The coefficient of determination is sensitive to the variability in data. As the amount of variability, or "scatter," around the fitted regression line increases, the value of r^2 decreases. An r^2 value of 1 means that all values fall on the regression line.

In some cases, an association between variables may exist, but it is not strictly linear. **Spearman's rank correlation coefficient (*r*)**, or **Spearman rho**, is the counterpart of the Pearson correlation coefficient for ordinal data. It is a **nonparametric measure** (see below) for use with data that are either reduced to ranks or collected in the form of ranks. It provides a way to quantify by how much two variables go up (or down) together without assuming that the relationship follows a straight line. The Spearman rho, like the Pearson correlation coefficient, yields a value from −1 to 1, and it is interpreted in the same way.

When an association between two variables is suspected, it is best to construct a **scatterplot** before deciding on an analysis strategy. A scatterplot may reveal unique patterns in the data such as outliers, clusters, and nonlinear relationships (or no apparent relationship at all) and may suggest not only the most appropriate analysis strategy but the **clinical relevance** of the suspected association.

EXAMPLE 9.4: CAN URINE COLOR IN DOGS BE USED TO ESTIMATE URINE-SPECIFIC GRAVITY?

Background: A key component of urinalysis is the assessment of urine-specific gravity (USG). Urine-specific gravity, as measured through refractometry, is typically ≥1.030 in dehydrated canine patients with normal renal function and can be used to distinguish between pre-renal and renal azotemia (abnormally high levels of nitrogen-containing compounds such as urea and creatinine in the blood). It is commonly assumed that a deeper yellow color of urine is indicative of more concentrated urine. However, there is little information available on the correlation between urine color (UC) and USG in canine patients and whether UC can be used to estimate USG. If a relationship exists, it may help veterinary personnel or owners assess hydration status in dogs.

Objectives: Cridge et al. (2018) conducted a study to determine (1) the degree of correlation between UC and USG, (2) if the use of a UC color chart would have an effect on the correlation between UC and USG, and (3) whether dark yellow (UC score of 4) could be used to predict whether the USG is ≥1.030.

Study Design: Cross-sectional.

Methods: One hundred medical records randomly selected from a pool of 1538 dogs that had undergone urinalysis at the Mississippi State University College of Veterinary Medicine (MSU-CVM) Teaching Hospital over the preceding 18 months, and that met study criteria, were evaluated. Urine color descriptors (clear, light yellow, yellow, dark yellow) were correlated with urine-specific gravity, as determined with a refractometer. Any urine samples that contained bilirubin crystals, significant proteinuria, glucosuria, or hematuria, factors that could alter urine color or specific gravity, were excluded from the study. A visual UC chart representative of the above four color descriptors was produced and used to score 95 canine urine samples received by the MSU-CVM clinical pathology laboratory over the subsequent 2 months. Clear urine was assigned a UC score of 1, light yellow a score of 2, yellow a score of 3, and dark yellow a UC score of 4.

Results: The Spearman rank correlation coefficient (r) for the relationship between UC and USG was positively correlated ($r = 0.44$), but not significantly improved by use of the UC chart ($r = 0.63$), due in part to the degree of scatter of individual data points. The degree of correlation was considered moderate. UC scores of 1 or 2 corresponded to median USG values below 1.030, and 80% of data points with a UC of 4 had a USG ≥1.030, indicative of adequate urine concentrating ability.

Conclusions and Significance: The results of the study suggest that UC can be used to estimate USG. If a dog's urine is clear (UC score of 1) or light yellow (UC score of 2) and the dog is clinically dehydrated, then there is likely inadequate urine concentrating ability. A UC dark-yellow urine score of 4 would suggest adequate urine concentrating ability in most (80%) patients.

FOLLOW-UP QUESTION 9.4

What are the limits on extrapolating the results of this study to clinical practice? See Answer 9.4 at the end of this chapter.

9.4 THE SELECTION OF AN APPROPRIATE STATISTICAL TEST

In most cases, statistical tests are used to estimate the probability of an alpha error, i.e., the likelihood of concluding that a difference exits when, in fact, it does not. The validity of each test depends on certain assumptions about the data. If the data at hand do not satisfy these assumptions, the resulting p_α may be misleading.

In research, there are many different statistical tests of significance (Shott, 2011; McDonald, 2014). Research studies differ in such things as the type of data collected, the kind of measurement used, and the number of groups used. These factors dictate which statistical test is appropriate for a particular study design.

For the uninitiated, the choice of an appropriate statistical test is not intuitively obvious. Shott (2011) developed flowcharts that summarize the reasoning process used to select among the most widely used statistical procedures in veterinary clinical research (Table 9.2). The flowcharts are intended to help readers understand and evaluate statistics in the veterinary literature and to help veterinary researchers select and interpret their statistics.

TABLE 9.2
Statistical Routines Commonly Used in Veterinary Clinical Research and Factors Influencing Their Selection

Censored data present
- Log-rank test
- Cox proportional hazards regression

Comparison of percentages
- X^2 test of association
- Fisher exact test
- McNemar test
- Cochran Q test
- X^2 test of hypothesized percentages

Comparison of two groups with respect to means or distributions
- Mann–Whitney test
- Paired and unpaired *t* test
- Separate- and pooled-variance *t* test
- Friedman test
- Paired sign test

Comparison of three groups with respect to means or distributions
- Kruskal–Wallis test
- 1-way ANOVA
- Friedman test
- 1-factor repeated measures ANOVA

Relationship between two variables
- Spearman correlation
- Pearson correlation
- Bivariate least squares regression
- X^2 test of association
- Fisher exact test

Relationship between dependent variable and independent variables
- Bivariate least-squares regression
- Multiple regression
- Logistic regression

Source: Shott S. *J Am Vet Med Assoc* 2011;239:948–950.

Among the questions asked at each branch of the flowcharts are (1) whether any of the data are censored, (2) the kinds of data being compared, (3) whether measurements are independent, and (4) whether the data are normally distributed. These issues are discussed further below.

> ***The validity of a statistical test depends on certain assumptions about the data. If the data at hand do not satisfy these assumptions, the resulting p_α may be misleading.***

9.4.1 Censoring

Was there a failure to completely follow up any individuals in the study? When outcome data are not available or only partially known for a subject at the time a study is completed, it is referred to as censored data. Censoring might occur if an individual withdraws from or is removed from the study

population, fails to provide all of the information requested, or if the study ends before the outcome being measured (survival time, time to disease recurrence) has occurred. The latter are sometimes referred to as **waiting times**. Censoring might also occur if a value lies outside of the range of a measuring instrument. If values are missing completely at random, the data sample is still likely to be representative of the population. But if the values are missing systematically, the analysis may be biased. When censored data are present, the number of statistical options is reduced (Shott, 2011). See Example 7.5 for an example of the use of censored data.

9.4.2 Level of Measurement

***What is the level of measurement: nominal, ordinal, or interval*? Nominal or categorical data** are used to categorize objects, individuals, conditions, etc. without ranking, such as breed, sex, or blood line. **Ordinal data** are ranked but do not fall on a uniform scale. Terms such as "light," "moderate," and "heavy" are used to describe ordinal data. **Interval or continuous data** are ranked on a scale of equal units, such as temperature, erythrocyte counts, etc. Refer to the section on scales in Chapter 2 for a further discussion and examples of each data type.

9.4.3 Number of Groups

***How many groups are there in the study: one, two, or more*?** If you want to find out whether a single group is representative of a specified population, then you are looking at **one group**. If you're interested in whether two samples come from the same population (the null hypothesis), then you are looking at **two groups**, whether they are two separate groups or the same group twice (as repeated measures over time). The same reasoning applies to **three or more groups**.

9.4.4 Nature of Groups

***What is the nature or character of your groups—independent or related*?** If the selection of an individual in one sample in no way influences the selection of an individual in another, then the groups are completely **independent**. In contrast, if groups have members that are "matched" or connected somehow to one another, then they are **related**.

Groups can be related when an individual serves as its own control, such as **repeated measures** conducted before and after treatment. Another way that groups can be related is when individuals are paired by characteristics such as age, breed, or sex before being randomly assigned to each group. Because of the prior matching, you would now have groups that are alike in age, breed, or sex. Any difference that emerges among groups could not be attributed to these three variables. Pairing is an example of adjusting for **covariance**, where the initial values for animals in each experimental group will influence subsequent values. Covariance is also of concern in multivariate analysis (see Chapter 6), where variables other than the one under consideration may influence the outcome.

9.4.5 Number of Categories

***How many categories are there*?** This question refers only to nominal data. The number of categories refers to the number of subdivisions that a group or sample is broken down into. For instance, the canine population of a veterinary hospital may be separated into four categories based on sex: male, female, male neutered, and female neutered.

9.4.6 Category Size

How many individuals or objects are in each of your categories? This question also refers only to nominal data.

9.4.7 Data

***How do you plan to use your data*?** This question primarily applies to ordinal data divided into two related groups. The data can be expressed in one of two forms: numbers (such as grade of heart murmurs) or as plus and minus signs (such as strength of immunodiagnostic test reactions).

9.5 PARAMETRIC AND NONPARAMETRIC TESTS

Statistical tests are referred to as either **parametric** or **nonparametric**. Many statistical tests for measurement variables assume that data are normally distributed (fit a bell-shaped curve). These are referred to as parametric data. Other data sets don't fit the normal distribution very well and must be analyzed with nonparametric tests. Parametric tests are more powerful than nonparametric tests, i.e., they have a higher probability of rejecting the null hypothesis when it should be rejected. Basic requirements for use of a parametric test are:

1. The groups in the samples are randomly drawn from the population.
2. The data are at the interval level of measurement.
3. The data are normally distributed.
4. The variances are equal.

Nonparametric tests have fewer and less stringent assumptions. Although they may meet the first requirement of parametric tests, they do not meet the rest. They are "distribution-free" tests whose level of measurement is generally nominal or ordinal. Nonparametric tests should also be used when sample sizes are very small, e.g., six or fewer.

9.6 SAMPLE SIZE

It is intuitively obvious that the more subjects that are entered into a study, the greater confidence we can have that differences among groups are not due to random variation. The question is, how many subjects are enough? One or more of the following variables must be considered to optimize the power of a particular study. These variables are: (1) the frequency of disease, (2) the amount of variability among individuals, (3) the difference in outcome between study groups, (4) p_α, and (5) p_b. Three common situations in which sample size must be considered follow.

> ***It is intuitively obvious that the more subjects that are entered into a study, the greater confidence we can have that differences among groups are not due to random variation. The question is, how many subjects are enough?***

9.6.1 Minimum Sample Size for Demonstrating an Extreme Outcome

The best example of this situation in veterinary medicine is when we have to decide how many animals to sample to determine whether a particular disease is present in the herd. This is a common concern in disease eradication or control programs. Here we only wish to detect the presence, rather than the prevalence, of disease in a herd. The type of error that we are trying to reduce is p_b, the likelihood of calling a herd negative when in fact it is positive (false-negative result).

Example: Returning to Example 9.2 above, consider a herd of deer in which 10% are CWD positive based on tonsil biopsy immunohistochemistry. If a tonsil biopsy is collected from one randomly selected animal in the herd, the probability that it will come from a CWD-free animal is 0.90. Thus, p_b is 0.90, i.e., we have a 90% chance of failing to detect CWD in the herd. If two animals

are sampled, then the chance that both samples were drawn from negative animals is 0.90 × 0.90, or 0.81.

Thus, the general formula for estimating p_b in the preceding example is

$$p_b = (1 - \text{Prevalence of disease})^n$$

where p_b = the chance that none of the sampled animals is harboring the disease and n = the sample size. This equation can be rearranged to estimate the required sample size for a given p_b

$$n_{inf} = \frac{\log(p_b)}{\log(1 - \text{Prevalence of disease})}$$

where n_{inf} = sample size for an **infinite population** (or very large relative to the sample size). If we set p_b at 0.05, then we would need to collect samples from approximately 29 animals to be 95% sure that at least one would be afflicted with CWD.

The astute reader will have noticed that the previous formula is true only for very large herd sizes. For example, if the deer herd consisted of 29 animals or less, and all were tested, we would be more than 95% certain of the presence or absence of disease. The sample size requirements for state and federal disease control programs are based on formulas that adjust for herd size. The sample size estimate will also depend on test sensitivity and specificity. Perhaps the most important factor in estimating sample size to detect the presence or absence of disease is the accuracy of our estimate of existing prevalence. Since the required sample size increases as estimated prevalence decreases, it is best to assume a "worst case" scenario, i.e., the lowest value for disease prevalence that we consider likely.

9.6.2 Minimum Sample Size for Estimating a Rate or Proportion with a Specified Degree of Precision

If we wish not only to detect disease, but also wish to estimate its prevalence, then a somewhat more complex calculation is needed to estimate sample size. As you might expect, the sample size is larger than that needed to detect only the presence of disease. Sample size for an **infinite population** (n_{inf}) is estimated by the formula

$$n_{inf} = \frac{(P)(1-P)Z^2}{d^2}$$

where P = the estimated prevalence of infection (as a decimal), Z corresponds to the degree of confidence in our estimate (usually Z = 1.96 for 95% confidence in our estimate), and d = the maximum difference between observed and true prevalence that we are willing to accept (as a decimal) (see http://www.macorr.com/sample-size-methodology.htm).

As before, sample size is inversely related to the amount of variability that we are willing to accept. Furthermore, test sensitivity and specificity, which are assumed to be 100% in this formula, will affect our estimate of the actual prevalence of the disease in the population.

To estimate the required sample size (n_{fin}) for estimating a rate or proportion when sampling from a **finite population** (N) the following conversion (see http://www.macorr.com/sample-size-methodology.htm) can be made:

$$n_{fin} = \frac{n_{inf}}{1 + (n_{inf} - 1)N}$$

9.6.3 Minimum Sample Size to Detect Differences among Groups in Studies of Risk, Prognosis, and Treatment

As indicated previously, a variety of statistical tests are available for determining the significance of outcomes in clinical studies. Corresponding sample sizes vary with the test being used. If the investigator is sure of which test will be used, then it is often useful to do "what if" experiments by "plugging in" some hypothetical results and seeing whether statistically significant differences could be detected. By trial and error, and a reasonable estimate of the range of possible outcomes, one can estimate the sample size that will be needed. The best approach is to discuss the proposed experimental design with a biomedical statistician before the study is conducted. This individual may suggest alternative designs and would most certainly be of aid in estimating the required sample size.

9.7 SAMPLING STRATEGIES

Ideally, an epidemiologic study should collect data from every individual in the **accessible population**, i.e., the population that is available for study. This may be possible when studying confined animal populations such as herds of cattle, stables of horses, etc. In other cases, the accessible population is too large or spread out over time, and a smaller sample of the population must be selected for study. Sampling should be conducted in such a way that the individuals selected for study are an unbiased representation of the population. Sampling strategies fall within two broad classes: **probability** and **nonprobability**, each with several versions (Hulley and Cummings, 2013). Examples of each are described below.

Regardless of the sampling strategy employed, several factors associated with the data collection process may influence the validity of results. This is especially true of questionnaire surveys, where the investigator is dependent on the willingness of sampled individuals to respond to the survey. The overall response rate has a direct effect on the power of a study, whereas bias in responders versus nonresponders may affect the validity of comparisons that are made. Finally, none of the sampling strategies described below ensure that the accessible population (for example, flea-infested dogs presented to a veterinary teaching hospital) is representative of the **target population** (all flea-infested dogs in the state, country, or world) to which the results will be generalized.

Given the variety of sampling options available, investigators should consult a biomedical statistician for advice on selecting the most appropriate sampling strategy.

9.7.1 Probability Sampling

Probability sampling uses a random process to assure that each member of a population has a specified chance of being selected. Probability sampling provides a scientific basis for saying that the intended sample represents the accessible population and for computing confidence intervals and statistical significance. Several versions of probability sampling follow.

9.7.1.1 Simple Random Sampling

In **simple random sampling**, every member of the population to be sampled is enumerated in a list (**sampling frame**), and then a subset is randomly selected for study. A table of random numbers may be used to select individuals. The representativeness of the resulting sample is dependent on the accuracy of the sampling frame and success in finding and enrolling the selected individuals.

9.7.1.2 Systematic Sampling

In **systematic sampling**, subjects are selected for study through a periodic process, such as every tenth individual in a list. This approach might be used for sampling a large herd of cattle at the time of processing through a chute, or poultry on the processing line in a packing plant. Systematic

sampling is technically a form of probability sampling, especially if the starting point is chosen at random. However, investigators should be alert for any natural periodicities in the population being sampled that might influence the representativeness of the sample.

9.7.1.3 Stratified Random Sampling

In **stratified random sampling**, the population is divided into subgroups according to characteristics such as age, breed, sex, or severity of clinical condition, and a random sample is taken from each of these **strata**. Stratified random sampling can be used to ensure consistency of precision across strata, or to ensure that geographically dispersed strata are proportionately represented. For example, in studying the incidence of adverse effects of early neutering, a feline population might be stratified by sex and age at gonadectomy (<5 and ≥ 5 months of age) and equal numbers of individuals then randomly selected from each stratum. This would yield incidence estimates for each sex/age at gonadectomy stratum with comparable precision. Alternatively, if we wish to estimate the regional prevalence of a disease among cattle, herd sampling could be proportional to the representation of dairy versus beef cattle in the entire population.

9.7.1.4 Cluster Sampling

Cluster sampling is the process of taking a random sample of natural groupings (**clusters**) of individuals from a population. Cluster sampling is useful for obtaining a representative sample from a widely dispersed population when it is impractical or costly to randomly sample the entire population. For example, a review of medical records of canine and feline cases of dental disease selected randomly from all cases seen in practices statewide would not be possible, as there is no statewide list of discharge diagnoses for private practices. The study could be conducted, however, by selecting a random sample of veterinary practices statewide and then reviewing all cases of canine and feline dental disease from each. **Two-stage cluster sampling** is used to draw a sample from populations that are organized into discrete subunits, such as census tracts or city blocks in human communities, or animal holding units in production facilities. The first stage consists of drawing a random sample of subunits for sampling. The second stage consists of drawing a random sample of individuals from the subunits selected in the first stage.

Cluster sampling provides a way to reduce the difficulty and expense associated with population-based sampling, but there are some disadvantages. As naturally occurring groups tend to be relatively homogeneous, a relatively large number of clusters, heterogeneous for the variables of interest, should be sampled to ensure that the sample is representative of the population. Furthermore, because of the way the sample is selected, data analysis is more complex than for the previously described sampling strategies.

EXAMPLE 9.5: WHAT ARE THE OBSTACLES TO RABIES CONTROL IN ENDEMIC REGIONS?

Background: Rabies is a preventable viral disease of mammals most often transmitted through the bite of a rabid animal. Globally, more than 15 million people receive rabies post-exposure prophylactic treatment annually, and an estimated 55,000 people die from rabies each year. Africa and Asia record the highest human rabies deaths worldwide, with an estimated 24,500 annual human deaths (Mucheru et al., 2014). Factors promoting dog vaccination, estimates of vaccine coverage, and knowledge about rabies are important for effective rabies control. These attributes are lacking in most countries recording high dog bite and rabies cases, including Kenya. Despite numerous government and private rabies vaccination campaigns, rabies remains endemic in some parts of Kenya due to inadequate coverage and high dog turnover rates.

Objectives: Mucheru et al. (2014) conducted a study to determine the rabies vaccination coverage among dogs at the household level and establish whether the level of knowledge on rabies disease influences dog vaccination practices in Kakamega County of Kenya.

Study Design: Cross-sectional.

Methods: The most recent census (2009) in Kakamega County of Kenya reported a total population of 1,660,651 residents and 398,709 households in the county, which covered an area of 3,244.9 square kilometers. A minimum sample size of 384 households (HH) was estimated based on an expected prevalence of 50% dog-owning HH (to achieve maximum sample size), a 95% CI, and desired accuracy of 5% (see http://www.macorr.com/sample-size-calculator.htm for a sample size calculator). A two-stage systematic random sampling strategy was followed to select study participants. In the first stage, 30 clusters were selected from the master frame of all Enumeration Areas (EAs; a geographic area canvassed by one census representative) using simple random sampling. Each cluster was the equivalent of one EA. The second stage consisted of the selection of households within clusters using systematic random sampling. As the number of households in each cluster varied, the sampling interval (N/n) for each cluster was determined by dividing the total number of households (N) in each cluster by the number of households to be interviewed for each cluster (13). At least 13 households and 7 dog-owning households selected per cluster were sampled. One member (above 18 years of age) from each household was interviewed with a structured questionnaire focusing on rabies knowledge and practice. A set of questions related to rabies knowledge and practice were used to score participant response. A score above the sample mean was equated to adequate knowledge and proper practices, respectively. An independent (unpaired) *t*-test was used to evaluate the differences of sample mean scores based on dog vaccination status. Bivariate analysis used dog vaccination status as the outcome variable to test the relationship between knowledge about rabies and practices.

Results: A total of 390 HH were enrolled, of which 338 owned one or more dogs, making up a population of 754 dogs. Only 35% (n = 119) of HH had dogs vaccinated within the past 12 months. There was a statistically significant difference in mean knowledge ($p < 0.01$) and practice ($p = 0.001$) of HH with vaccinated dogs compared to ones with unvaccinated dogs (Table 9.3). Participants with adequate rabies knowledge were more likely to have proper health-seeking practices and proper handling practices of suspected rabid dogs.

TABLE 9.3
Summary of Factors Independently and Significantly Associated with Having a Vaccinated Dog in Households on Multivariate Analysis, Kakamega County, Kenya, 2013

Variable	Odds Ratio	*p* Value
Having formal employment	2.704	0.006
Having food prepared specifically for dog in household	2.49	0.0006
Knowing of at least two annual vaccination clinics	3.51	0.000007
Household had dog implicated to have bitten someone	2.96	0.0006
Knowing of a disease called rabies	3.13	0.003
First rabies vaccination done when dog is ≤1 yr old	2.64	0.0006
Knowing exposure to rabies without treatment leads to death	2.4	0.00027

Source: Mucheru GM et al. *Pan Afr Med J* 2014;19:255. With permission.

Conclusions and Significance: The authors concluded that the rabies vaccination level (herd immunity; see Chapter 12) was below the 80% recommended to control rabies outbreaks.

FOLLOW-UP QUESTION 9.5

What steps could be taken to raise the level of herd immunity in this population? See Answer 9.5 at the end of this chapter.

EpiTools (Sergeant, 2019), available at http://epitools.ausvet.com.au/, includes a collection of sampling strategies and statistical routines that can be used to estimate disease prevalence or demonstrate freedom from disease through structured surveys, or in other epidemiological applications.

9.7.2 Nonprobability Sampling

In some cases, a nonprobability sampling design may be the only option available to the investigator. Reasons include cost, convenience, and the nature of the accessible population (those willing to submit data, for example). If a nonprobability sample is to be used, it is important that it approximate, as closely as possible, the kind of sample that would be obtained by probability sampling, as statistical tests are likely to be applied to the results. This is the same consideration when choosing an accessible population, i.e., that it be representative of the target population.

9.7.2.1 Consecutive Sampling

Consecutive sampling involves taking every patient from the accessible population who meets the selection criteria over a specified interval or number of patients. If the data to be gathered can be influenced by temporal disease patterns, then the sampling period should be of sufficient duration to accommodate this variation.

9.7.2.2 Convenience Sampling

Convenience sampling is the process of selecting those members of the accessible population who are easily accessible. Patient selection is often based on willingness of owners to participate in the study. As such, there is always the risk that study subjects do not accurately represent the population at large. Investigators should address this concern when discussing study results.

9.7.2.3 Judgmental Sampling

Judgmental sampling involves selecting from the accessible population those individuals judged most appropriate for the study. In this regard, judgmental sampling is susceptible to the same biases as convenience sampling.

Medical records data can be used by veterinary practitioners to better understand and anticipate health problems of importance in cats and dogs they examine and to better communicate with clients regarding the most prevalent disorders. Observational studies in companion animal research are often based on patients seen at veterinary medical teaching hospitals. Since many of these animals are referred to the VMTH with diseases that are difficult to diagnose and treat, they may not be representative of the general population seen at private veterinary practices. The generalizability of results may be limited by a number of additional factors, including:

- The representativeness of participating practitioners among all practitioners
- The lack of case definitions and possibility of disease misclassification and underreporting of disorders requiring extensive or expensive diagnostic testing

- Underreporting due to inconsistency in recording diagnostic codes
- Failure to distinguish between new (incident) and existing (prevalent) disorders that may limit the use of data for monitoring disease trends

9.8 MULTIPLE COMPARISONS

Some studies, referred to as **hypothesis testing**, are designed to evaluate the effect of one variable (such as a risk factor, prognostic factor, or treatment) upon an outcome. However, during the course of a study in which statistically significant results are found, it is often tempting to break groups down into smaller groups to search for additional associations. This process is referred to as **hypothesis generation** (or more disparagingly as *data dredging* or a *fishing expedition*).

One problem with such multiple comparisons is that the resulting subgroups contain fewer individuals than did the initial groupings. Consequently, the number of individuals in these groups may be too small to allow statistically significant differences to be detected. A second problem in making multiple comparisons is similar to the problem encountered in parallel testing—if enough comparisons are made, the investigator is more likely to detect at least one that will be statistically significant, irrespective of the true state of affairs. Consequently, results derived from multiple comparisons should be considered hypotheses to be tested in follow-up studies.

If enough comparisons are made, the investigator is more likely to detect at least one that will be statistically significant, irrespective of the true state of affairs.

REFERENCES

Christley RM and Reid SW. No significant difference: Use of statistical methods for testing equivalence in clinical veterinary literature. *J Am Vet Med Assoc* 2003;222:433–437.

Cridge H, Wills RW, and Lathan P. Correlation between urine color and urine specific gravity in dogs: Can urine color be used to identify concentrated urine? *Can Vet J* 2018;59:178–180. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5764211/

Hazuchova K, Held S, and Neiger R. Usefulness of acute phase proteins in differentiating between feline infectious peritonitis and other diseases in cats with body cavity effusions. *J Feline Med Surg* 2017;19:809–816.

Hulley SB and Cummings SR. *Designing Clinical Research*. Philadelphia: Wolters Kluwer/Lippincott Williams & Wilkins; 2013.

Marchesi MC, Timpano, CC, Busechian S et al. The concordance between endoscopic and histological diagnosis in 114 dogs affected by gastric disease. *Veterinaria Italiana* 2017;53:309–313. Full text available at: http://www.izs.it/vet_italiana/2017/53_4/309.htm

McDonald JH. *Handbook of Biological Statistics*, 3rd ed. Baltimore: Sparky House Publishing, 2014. Full text available at: http://www.biostathandbook.com/index.html

McHugh ML. Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)* 2012;22:276–282. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/

Mucheru GM, Kikuvi GM, and Amwayi SA. Knowledge and practices towards rabies and determinants of dog rabies vaccination in households: A cross sectional study in an area with high dog bite incidents in Kakamega County, Kenya, 2013. *Pan Afr Med J* 2014;19:255. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4382068/

Saverio P, Alessia G, Vito T et al. Critical assessment of the diagnostic value of feline α_1-acid glycoprotein for feline infectious peritonitis using the likelihood ratios approach. *J Vet Diagn Invest* 2007;19:266–272. Full text available at: https://journals.sagepub.com/doi/full/10.1177/104063870701900306

Sergeant ESG. 2019. *Epitools Epidemiological Calculators*. Ausvet Pty Ltd. Available at: http://epitools.ausvet.com.au

Shott S. Wrapping it all up. Statistics in veterinary medicine. Flowcharts. *J Am Vet Med Assoc* 2011;239:948–950.

Swift BMC, Bennett M, Waller K et al. Anthropogenic environmental drivers of antimicrobial resistance in wildlife. *Sci Total Environ* 2019;649:12–20. Full text available at: https://www.sciencedirect.com/science/article/pii/S0048969718331449?via%3Dihub

Trefz FM, Constable PD, and Lorenz I. Effect of intravenous small-volume hypertonic sodium bicarbonate, sodium chloride, and glucose solutions in decreasing plasma potassium concentration in hyperkalemic neonatal calves with diarrhea. *J Vet Intern Med* 2017;31:907–921. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5435070/

Wolfe L, Watry MK, Sirochman MA et al. Evaluation of a test and cull strategy for reducing prevalence of chronic wasting disease in mule deer (*Odocoileus hemionus*). *J Wildl Dis* 2018;54:511–519. Full text available at: https://bioone.org/journals/journal-of-wildlife-diseases/volume-54/issue-3/2018-01-015/EVALUATION-OF-A-TEST-AND-CULL-STRATEGY-FOR-REDUCING-PREVALENCE/10.7589/2018-01-015.full

Zwillinger D and Kokoska S. *CRC Standard Probability and Statistics Tables and Formulae*. Hoboken, NJ: Taylor & Francis Group; 2000.

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 9.1: Receiver-operating characteristic analysis (see Chapter 3) could be used to assess the likelihood of FIP based on the level of an APP measured in serum or effusion and determine the best cutoff value. ROC analysis was, in fact, performed by the authors (Figure 9.3). AGP in the effusion was shown to be the best marker to distinguish between cats with and without FIP; a cutoff value of 1550 μg/mL had a sensitivity and specificity of 93% each for diagnosing FIP. The cut-off values for the tested parameters were chosen preferably to obtain a high specificity because a false-positive result could be potentially fatal for the cat.

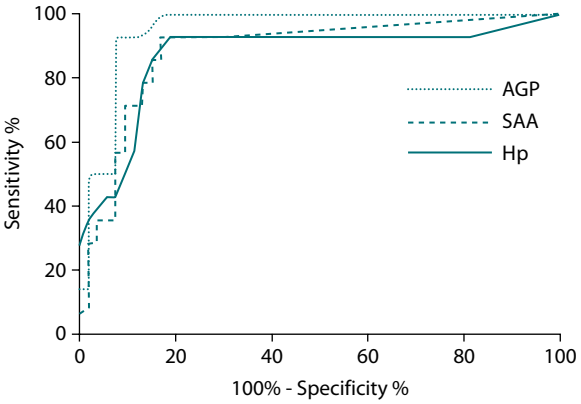

FIGURE 9.3 Receiver operating characteristic curves of the three acute phase proteins haptoglobin (Hp), α1-acid glycoprotein (AGP), and serum amyloid A (SAA) in the effusion of cats with feline infectious peritonitis (FIP; n = 14) and cats without FIP (n = 53). (From Hazuchova K et al. *J Feline Med Surg* 2017;19:809–816. With permission.)

Answer 9.2: Increase sample size. The confidence limits for a proportion estimated by all three techniques described above (normal approximation, binomial lookup tables, and online calculator) are greatly influenced by sample size. As sample size increases, the confidence limits become tighter around the mean. If, for example, the sample size for CWD prevalence determination were increased 10-fold from 66 to 660, the 95% binomial confidence limits would be 0.0309 to 0.0643 (versus 0.0095 to 0.1271 for a sample size of 66) at the same prevalence of 0.0455 (4.55%).

Answer 9.3: Positive and negative predictive values. For any combination of test sensitivity and specificity, the predictive value of a positive or negative test varies with the prevalence of likelihood of disease. See Chapter 3, "Evaluation of Diagnostic Test" for further information.

Answer 9.4: Caution should be exercised when extrapolating the results of this study to clinical patients, as this study included only patients with no evidence of abnormalities on urinalysis. This involved the exclusion of samples with bilirubinuria or hematuria, which would make the urine a darker color. The authors suggest that further research is required to determine the effect of co-morbidities on the correlation between UC and USG in canine patients.

Answer 9.5: Raise the level of herd immunity through a mass vaccination campaign coupled with more innovative ways of translating knowledge into proper rabies control practice by dog owners.

10 Medical Ecology and Outbreak Investigation

10.1 INTRODUCTION

The previous chapters have focused on clinical epidemiology and the role of population characteristics in veterinary decision-making. We have discussed the criteria by which clinically normal findings are distinguished from abnormal findings; factors affecting the interpretation and use of diagnostic tests; ways to measure the frequency of clinical events and their use to assess risk, prognosis, and treatment outcomes; and the role of chance in clinical research. In the following chapters, we will discuss the dynamics of disease in populations, i.e., **medical ecology**. We will also learn how to conduct an outbreak investigation using all of the concepts, tools, and approaches discussed in previous chapters.

One of the things that distinguishes veterinary from human medicine is that veterinarians are frequently called upon to diagnose and treat disease in populations as well as individuals. The health of an individual animal may be less important than that of the flock, kennel, or herd. However, the disease status of an individual animal frequently reflects that of the population from which it came. In other words, the animals that we see as clinicians may be regarded as **sentinels** for disease in the population.

> ***The disease status of an individual animal frequently reflects that of the population from which it came … the animals that we see as clinicians may be regarded as sentinels for disease in the population.***

Practitioners are frequently called upon to participate in local, state, and federal disease control programs. To perform in this capacity, veterinarians must understand and be able to communicate the scientific basis of these disease control programs to their clients. As veterinarians, we are expected to know how diseases are introduced, spread, and persist in animal populations. We must determine the cause of disease and also devise a plan to reduce disease frequency to an "acceptable" level. What is acceptable will depend on the cost of the disease and the cost of control.

10.2 ISSUES IN THE EPIDEMIOLOGY OF A DISEASE

A number of issues emerge when considering the epidemiology of any disease. A distinction must be drawn between the **life cycle** of a disease agent, which describes the movement of a disease agent in the environment, and the **epidemiology of a disease** (or medical ecology), which describes the dynamics of a disease in the population. The life cycle of the disease agent is only part of the story. The major issues in the epidemiology of a disease are described below.

10.2.1 Occurrence

In Chapter 5, some of the measures of disease frequency were discussed. **Occurrence** refers to the frequency distribution of disease over space (spatial or geographic occurrence), time (temporal occurrence), or within

a host population (demographics). This information is useful not only to gain a better appreciation of the significance of the disease, but also its probable cause, source, and mode of transmission.

10.2.2 Cause

Causes, or **determinants**, of disease include the etiologic agents directly responsible for disease and other factors that facilitate exposure, multiplication, and spread in the population. Disease determinants can be categorized as **agent, host, and environment** (or management) factors.

> ***Disease determinants can be categorized as agent, host, and environment (or management) factors.***

10.2.3 Susceptibility

Host determinants of disease occurrence include both individual characteristics of hosts that render them susceptible or resistant to disease, and population characteristics, such as the level of **herd immunity**. Just as parasitic organisms have defined life cycle stages, a diseased population may be divided into **epidemiologic classes**. Typical epidemiologic classes are susceptible, incubating, sick, recovered, and immune. The proportion of the population in each of these classes will determine, in part, the dynamics of disease transmission within the population.

10.2.4 Source

Sources of disease agents include (1) recently infected individuals, (2) carrier animals (animals with inapparent infections that are also transmitters or potential transmitters of the infectious agent), (3) intermediate hosts and vectors, and (4) the environment. For every clinical case of a disease, there may be numerous other inapparent infections. Some may be individuals in the incubation or prepatent phase of the disease. Others may be recovered individuals who continue to harbor the organism. If these individuals are also infectious, they may be a major source, or reservoir, of infection for susceptibles.

> ***A diseased population may be divided into epidemiologic classes. Typical epidemiologic classes are susceptibles, incubating, sick, recovered, and immune.***

10.2.5 Transmission

Diseases are broadly classified as **transmissible** or **non-transmissible**. Within these two broad categories there are a number of specific modes of transmission. A distinction must be made between the **mode of transmission** and the **route of infection**. It would be incorrect to say that the mode of transmission is via the respiratory tract since we have not indicated whether the organisms gained access via droplet transmission (direct transmission), droplet nuclei, or dust (airborne transmission). The respiratory tract is really a route of infection rather than a mode of transmission.

10.2.6 Cost

In food-producing and other animals raised and managed for profit, the impact of disease is frequently described in terms of performance or economics, rather than morbidity and mortality. Likewise,

decisions as to whether to treat or cull the animal may be determined in large part by economics. Any assessment of cost should include the cost of disease control.

10.2.7 Control

Ultimately, the practitioner must devise a plan for the reduction of disease risk or frequency in the population. This may be accomplished through disease prevention, control (treatment), or eradication.

10.3 OUTBREAK INVESTIGATION

Outbreak investigation, sometimes referred to as "field epidemiology," is similar, in principle, to examination of a patient in a hospital setting. In both instances, history and physical and laboratory examinations are used to try to identify the cause(s) of disease at the individual or herd level. Working hypotheses at the herd level are (1) diseases usually have multiple causes, and (2) disease events are not randomly distributed in a population. Typically, disease frequency and distribution data are collected and analyzed to identify disease patterns (occurrence), which are then analyzed to suggest determinants of disease.

By tracing the steps involved in an outbreak investigation, we can better appreciate the importance of the issues in the epidemiology of a disease. The steps are analogous to the systematic approach (**SOAP**—see Chapter 2) used with individual patients. The following steps in an outbreak investigation have been adapted to a veterinary context from a U.S. Centers for Disease Control self-study course (Dicker et al., 2012).

10.3.1 Descriptive Phase (Subjective, Objective Data)

The distribution of cases during an outbreak follows certain patterns in time (chronology), space (geography), and hosts (demography). The chronological distribution of disease events can be recognized by plotting the frequency of new cases over time, resulting in an epidemic curve. The geographic distribution can be recognized using various types of maps, most commonly spot maps. The demographic patterns of disease distribution can be identified by comparing frequency rates in different strata based on age, sex, breed, etc., and depicted as attack rate tables or graphs. Among the questions asked during this phase of outbreak investigation are the following:

1. What are the characteristics of the clinical syndrome, e.g., the case definition?
   a. What signs were/are observed in live and dead animals?
   b. What was the incubation period?
   c. How long did signs last?
   d. What is the prognosis for diseased animals?
2. What are the temporal, spatial, and demographic patterns of disease?
   a. When did the cases occur?
   b. Where did the cases occur?
   c. What was the incidence of disease, e.g., how many animals were at risk and how many were affected?
   d. What are the characteristics of the affected and unaffected animals?
   e. How rapidly did the disease spread and what is the likely mode of transmission?
   f. Are any other domestic animals or wildlife affected; is there any concurrent human illness?
3. What is the herd history?
   a. Describe the management and husbandry practices, including housing, feed, water.
   b. Describe disease control/hygiene practices including vaccination, parasiticides/dewormers, other treatments, vermin and pest control, and waste disposal.

   c. Describe the herd's production/disease history.
   d. Has there been contact with other domestic animals or wildlife?
   e. Has there been any animal movement or introductions recently?
   f. Have there been any health problems in adjacent herds?
4. What is the environmental history?
   a. What has the weather been like?
   b. Describe the geographic location, e.g., topography, soil type, vegetation.
   c. Have fertilizers, herbicides, or pesticides been used recently?

The answers to the above questions should guide the formulation of **testable hypotheses** as to the source, identity, and mode of transmission of the etiologic agent. Sample collection and the choice of appropriate diagnostic test procedures follow.

10.3.2 Analytic Phase (Assessment)

During this phase, the plausibility of the above hypotheses is evaluated. The descriptive data are compared and analyzed in light of what is known about diseases on the differential list and whatever laboratory test results had been requested.

1. What associations exist, e.g., what risk factors appear to be associated with the disease?
2. What is the probable source of the etiologic agent and how is it being spread?
3. What is the probable cause of the disease?
4. How much does the disease cost?

10.3.3 Intervention (Plan)

What are you going to do? This is why you became involved in the first place.

1. Are current measures adequate to control the outbreak? What else should be done?
2. What immediate and long-term preventive options are available?
3. What are the economic benefits/consequences of these options?

In the following chapters, each of the issues in the epidemiology of a disease is discussed. Examples have been chosen to illustrate how outbreak investigations are pursued.

REFERENCE

Dicker RC, Coronado F, Koo D et al. *Principles of Epidemiology in Public Health Practice*, 3rd ed. Atlanta: Centers for Disease Control and Prevention (CDC); 2012. Full text available at: https://www.cdc.gov/csels/dsepd/ss1978/SS1978.pdf. Outbreak investigation module available online at: https://www.cdc.gov/csels/dsepd/ss1978/lesson6/section2.html

11 Measuring and Expressing Occurrence

11.1 INTRODUCTION

Earlier in the text, we discussed frequency of clinical findings and disease and made a distinction between incidence and prevalence. Occurrence refers to the frequency distribution of disease over space (spatial or geographic occurrence), time (temporal occurrence), or both (spatiotemporal occurrence) within a host population. This information is useful not only to gain a better appreciation of the significance of the disease, but may also suggest its probable cause, source, and mode of transmission.

11.2 CASE DEFINITION

The first step in any disease investigation is identification of the cases and noncases. This is not as easy as it might first appear. In studies of the characteristics of experimentally induced disease, animals are easily separated into cases and noncases on the basis of their exposure history. When faced with a disease outbreak, however, we don't usually know the nature of the exposure or which animals were exposed. We only have our perceptions of which animals are sick and which are not.

11.2.1 Based on Disease Signs, Symptoms, and Epidemiology

Cases may be defined on the basis of a discrete set of signs and symptoms. However, few animals show the complete range of disease signs, and minimal criteria for a diagnosis often have to be established. Biological variation among true cases and noncases has the effect of including cases among the noncases and vice versa. Furthermore, in any population there will always be animals with inapparent infections. Some cases will be incorrectly assigned to the noncase group. Clinical signs alone are seldom restrictive enough to exclude animals who are not suffering from the disease in question, but who may exhibit signs consistent with it. In these cases, epidemiologic criteria, such as the occurrence of the disease (see below), may be added to the case definition.

11.2.2 Based on Performance

Cases do not have to be defined on the basis of a clinically defined syndrome. Frequently we are interested in identifying risk factors associated with substandard performance. Producers usually become aware of a disease condition by its adverse effect on animal performance.

11.3 REPORTING DISEASE OCCURRENCE

The occurrence of disease in a population may be reported in three different ways:

1. *Host characteristics*, such as age, sex, and breed;
2. *Time*, which includes date of onset; or
3. *Place*, from within a housing unit to geographic distribution.

Scrutiny of the results of such classification enables one to recognize characteristics common among affected individuals, and rare among the healthy (Dicker et al., 2012).

11.3.1 Host Distribution

11.3.1.1 Attack Rate

Earlier in this book, we discussed incidence and prevalence, incidence being the number of new cases occurring in a susceptible population over a defined time interval, and prevalence being the number of sick individuals at any given point in time. A third rate that is frequently used, particularly during outbreak investigations, is the **attack rate**. An attack rate measures the proportion of the population that develops disease during a specified period, such as the duration of an outbreak, among the total exposed at the beginning of the outbreak (Dicker et al., 2012). The attack rate equals

$$\frac{\text{Number who become sick}}{\text{Number at risk at beginning of outbreak}}$$

The attack rate is essentially an incidence rate where the time period of interest is the duration of the epidemic. For analytical purposes, attack rates may be subdivided based on exposure status.

11.3.1.2 Crude versus Adjusted Rates

Comparison of disease rates among different groups is fundamental to determining the cause, source, and probable mode of transmission of a disease. Since comparison of crude rates (see Chapter 5) can lead to erroneous conclusions, it is necessary to adjust for any host factors that might interfere with an accurate comparison. Rates are commonly adjusted for age, breed, and sex (see Chapter 5).

11.3.2 Temporal Distribution

Most diseases have characteristic patterns of temporal occurrence. When disease is first recognized in a population, frequency data should be used to construct an **epidemic curve**. An epidemic curve gives a convenient pictorial depiction of the epidemic, and certain limited deductions may be drawn. Specifically, we want to know whether the disease is sporadic, endemic, or epidemic. The answer to this question often gives important clues as to the mode of transmission of a disease agent and its identity, and suggests what subsequent steps should be taken.

11.3.2.1 Sporadic Disease

A disease is **sporadic** when it occurs rarely and without regularity in a population unit. A sporadic pattern of occurrence elicits the question: "Where is the disease when it apparently is not around?" One explanation might be that infection exists in the population inapparently and only in occasional animals do signs of disease evidence themselves. An example might be fleabite dermatitis in cats and dogs. Most have fleas, but few develop severe reactions to infestation. A second explanation might be that the infection is generally absent, and the disease is noted only when it is introduced into the population with an infected animal (as bovine tuberculosis), a suitable vector (as West Nile virus), or occasional contact with an environmental source, either animal (as plague) or inanimate (as tetanus).

11.3.2.2 Endemic Disease

A disease is **endemic** when it occurs with predictable regularity in a population with only minor fluctuations in frequency pattern over time. A disease may be endemic at any level of occurrence, but the term **hyperendemic** is often used when a high proportion of animals are affected within a given geographic area or population. Herd infestations with internal parasites tend to occur as endemic diseases.

11.3.2.3 Epidemic Disease (Outbreak)

A disease is **epidemic** when its frequency within the population during a given time interval is clearly in excess of its expected frequency. The epidemic occurrence of disease is not based on absolute numbers or rates; it is a purely relative term. Thus, whether an observed frequency of any

particular disease constitutes an epidemic would vary from one place and population to another. An epidemic implies a clustering of disease in space as well as time. **Outbreak** is a somewhat less precise term, roughly synonymous with epidemic. The shape of the epidemic curve may provide clues as to the nature of exposure of susceptible individuals and the etiologic agent. A **point (or common) source epidemic** will typically have an epidemic curve that is skewed to the right; exposure occurs over a relative short period of time, and the *tail* reflects variable incubation periods. **Propagating epidemics** often have a curve skewed to the left, reflecting long-term exposure to the disease agent, generally reflecting transmission among individuals rather than from a common source. A **pandemic** is a large-scale epidemic over a wide geographic region, usually affecting a substantial proportion of the population. Conditions leading to an epidemic are essentially the same as those outlined for sporadic disease. Whether a disease presents as sporadic or epidemic is also a function of the efficiency of transmission of infection from infected to susceptible animals. Examples of epidemic curves are depicted in Figures 11.1 and 11.2 and described in the following two examples.

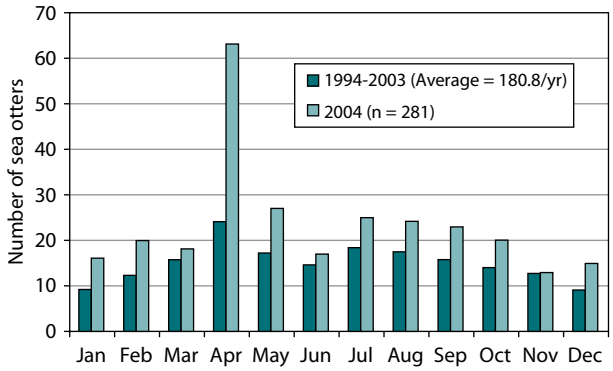

FIGURE 11.1 Number of southern sea otters stranding range wide by month—Estero Bay and Morro Bay along the central coast of California. (From Miller MA et al. *Vet Parasitol* 2010;172:183–194. With permission.)

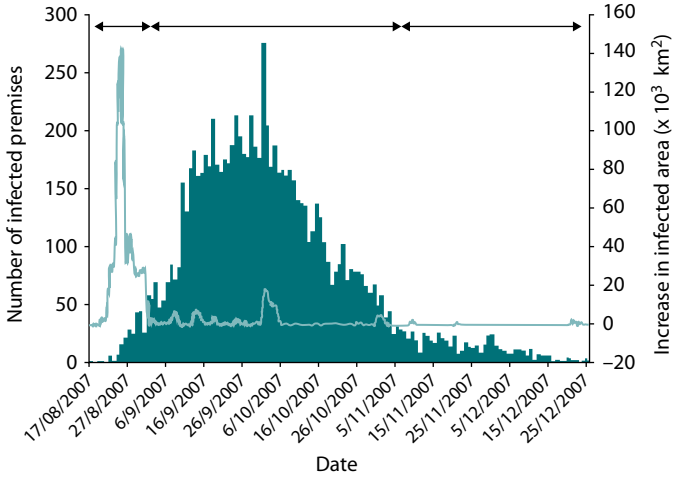

FIGURE 11.2 The spatio-temporal epidemic curve for the Australian equine influenza epidemic. The histogram represents the number of new infections by estimated first clinical signs each day, while the frequency polygon represents the increase in the total area of infected land every 2 days. Each phase of the epidemic is represented by an arrow (with arrows from left to right indicating dispersal, local spread, and epidemic fade-out phases). (From Cowled B et al. *Prev Vet Med* 2009;92:60–70. With permission.)

EXAMPLE 11.1: WHAT CAUSED AN OUTBREAK OF SEA OTTER MORTALITY IN CALIFORNIA?

Background: The southern sea otter (*Enhydra lutris nereis*) is a federally listed threatened marine mammal. Strandings (dead, sick, or injured sea otters) typically peak in the spring each year in California, with the highest shoreline deposition of live and dead otters occurring during April. Between 1994 and 2003, the mean number of otters recovered during April was 24 animals. In contrast, during April 2004, 63 sick and dead sea otters were found over 18 km of coastline near Morro Bay, California, >2.6 times the average for April over the preceding decade. This was the highest monthly deposition of live and dead otters ever recorded during 30 years of stranding data collection in California.

Objectives: Miller et al. (2010) conducted an outbreak investigation to identify the etiologic agent(s) responsible for sea otter morbidity and mortality and identify factors that contributed to the outbreak.

Study Design: Cross-sectional.

Methods: Because of the point source nature of the event and clinical signs consistent with severe, acute neurological disease, exposure to a chemical or marine toxin was initially considered. A detailed investigation was conducted that included (1) postmortem examinations and histopathology; (2) tissue and serum sample testing for bacteria, fungi, viruses, and biotoxins; (3) testing of marine invertebrates consumed by southern sea otters for biotoxins; and (4) sample testing for protozoa by cell culture, serology, immunochemistry, and DNA sequence analysis. Observations regarding local sea otter abundance, foraging ecology, and prey selection were collected by state and federal biologists.

Results: Detailed postmortem examinations revealed lesions consistent with an infectious etiology, and further investigation confirmed the protozoan parasite *Sarcocystis neurona* as the underlying cause through PCR-positive test results from tissue samples and detection of high titers of anti-*S. neurona* IgM and IgG in serum of stranded animals. Evidence to support the point source character of this outbreak include the striking spatial and temporal clustering of cases (Figure 11.1). Concurrent exposure to the marine biotoxin domoic acid may have enhanced susceptibility of affected otters to *S. neurona* and exacerbated the neurological signs exhibited by stranded animals. Other factors that may have contributed to the severity of this outbreak include a large rainstorm that preceded the event and an abundance of razor clams (*Siliqua patula*), marine invertebrates consumed by southern sea otters, near local beaches attracting numerous otters close to shore within the affected area and contributing to increased risk of contagious spread of *S. neurona*.

Conclusions and Significance: The authors propose that the otters dying due to *S. neurona* infection fed on a common food source (razor clams) near the shoreline, thereby enhancing their exposure to *S. neurona* sporocysts released from coastal lagoons during the large storm event. The point source character of this event, the striking consistency of lesions, and detection of high concentrations of anti-*S. neurona* antibodies in affected animals is unprecedented and suggests a point source exposure to *S. neurona*, most likely as infective sporocysts from feces of the terrestrial definitive host, the opossum (*Didelphis virginiana*). This is the first report of a localized outbreak in marine wildlife caused by apicomplexan protozoa and demonstrates the importance of land–sea transport as a potential source of human and animal exposure to biological pathogens.

FOLLOW-UP QUESTION 11.1

What measures might be taken to reduce the likelihood of a similar event in the future? See Answer 11.1 at the end of this chapter.

EXAMPLE 11.2: HOW DID THE AUSTRALIAN EQUINE INFLUENZA EPIDEMIC OF 2007 SPREAD AND HOW WAS IT BROUGHT UNDER CONTROL?

Background: Equine influenza (EI) has a worldwide distribution and is caused by an influenza A virus (family *Orthomyxoviridae*). Infection of equidae is characterized by respiratory disease. In August 2007, Australia—which had previously been free of EI—experienced a large outbreak that was associated with imported horses. Despite a rapid and effective eradication campaign that limited the spread of the infection to just two eastern states (New South Wales [NSW] and Queensland [Qld]), nearly 10,000 premises were infected during the epidemic. The resulting program to eradicate EI was a multifaceted coordinated national response involving cooperation between Commonwealth (national) and State governments and industry. These measures contained and ultimately eradicated the epidemic, but it is important to determine which control tools were the most effective.

Objectives: Cowled et al. (2009) used spatial and temporal analytical techniques to describe the epidemic, quantify important descriptors, and generate hypotheses about how the propagating epidemic progressed and which control tools assisted in eradication.

Study Design: Cross-sectional.

Methods: The investigators first focused on describing the epidemic spatially and temporally. Three datasets were used in this analysis: a dataset of IPs, a dataset of vaccinated premises, and a dataset of the horse population at risk. These datasets were created from data supplied by the NSW and Qld governments. Premises were recorded as infected if they had clinical evidence of infection and laboratory testing-confirmed disease. An epidemic curve was created by plotting the number of premises on which equids first displayed clinical signs for each day of the epidemic on the Y axis, and day of the epidemic on the X axis. The change in the size of the infected area over time was also calculated and depicted graphically. Cumulative incidence and incidence rates (per animal year at risk) were calculated using the number of IPs (rather than individual horses) as the case units. Cumulative incidence was expressed as the proportion of all premises infected for the entire epidemic. Similarly, incidence rates per animal were estimated from the number of IPs, population at risk, and epidemic length.

Results: The epidemic was assumed to have begun with the first reported clinical signs of EI on August 17, 2007, and was assumed to have ended 130 days later with the first reported clinical signs on the last premises in Queensland reported to be infected on December 25, 2007. The epidemic curve peaked on October 1, 2007 (276 IPs). Three phases in the spatio-temporal epidemic curve were apparent, which the authors referred to as dispersal, local spread, and epidemic fade-out phases (Figure 11.2). The **dispersal phase** extended from the start of the epidemic until the initial spatial spread of infection ended, as indicated by the flattening of the spatial polygon. This phase was characterized by relatively few new IPs, but a very large expansion in the area of land infected, largely associated with movement of infected horses before implementation of the response program. The **local spread phase** extended from the end of the dispersal phase until the completion of the descending part of the epidemic curve. This phase was characterized by a large increase in the number of new IPs, with little increase in the area of infected land. The **epidemic fade-out phase** included that part of the epidemic curve representing waning and secondary peaks. It was characterized by low numbers of new cases at a gradually declining incidence rate. Movement restrictions to limit spatial dissemination and seeding of susceptible populations of equines were implemented August 25, 2007, and spatial dispersal of infection rapidly declined immediately afterward. Vaccination began on September 29, 2007, but the first round of vaccinations (i.e., the first vaccination in the immunization course) was not substantially complete until November 15, 2007.

Conclusions and Significance: Post-outbreak analysis of epidemics is necessary to understand how diseases spread, thus leading to improved contingency plans for managing future outbreaks. In particular, analysis of data from the Australian EI epidemic offered a rare opportunity to explore the spread of a highly contagious disease in a fully susceptible population. The authors suggest that movement restrictions contributed more to the control of the Australian EI epidemic than did vaccination.

FOLLOW-UP QUESTION 11.2

What evidence might support the authors' conclusion that movement restrictions played a greater role in controlling the epidemic than did vaccination? See Answer 11.2 at the end of this chapter.

11.3.3 Time Series Analysis

Time series analysis focuses on the detection, description, and measurement of patterns or periodicities from temporal disease occurrence data (Schwabe et al., 1977; Zhang et al., 2014). The purpose of time series analysis is to identify periods of high or low risk so that causal associations can be explored.

Patterns of disease occurrence may reflect the influence of one or more of the following: (1) long-term (secular) trend, (2) seasonal fluctuation, (3) residuals (cyclic and irregular variation).

Decomposition methods are among the most traditional methods used to conduct a time series analysis (Hartnack et al., 2009; Zhang et al., 2014; Molina et al., 2018). These methods can be used to break down the original time series into its component parts. Because decomposition time series methods do not involve a lot of mathematics or statistics, they are relatively easy to explain to the end user. This is a major advantage because if the end user has recognition of how a forecast was developed, he or she may have more confidence in its use for decision-making.

The decomposition model assumes that patterns of disease occurrence (incidence or prevalence) reflect the influence of one or more of the following: (1) long-term (secular) trend, (2) seasonal fluctuation, and (3) residuals (cyclic and irregular variation, or error). **Secular trends** are overall rises or declines in disease frequency that occur gradually over long periods of time. A trend can be identified from time series data by (1) visual observation of plotted raw data, (2) the moving average method, (3) exponential smoothing, or (4) least-squares regression. A **moving average** is a series of data averages centered at each successive measurement point on the time scale (Zhang et al., 2014).

Seasonal fluctuations are regular changes in disease frequency with periods shorter than a year that are usually related to known factors such as rainfall, month of the year, temperature, timing of production events, etc. Three-month moving averages help smooth out short-term data fluctuations and may approximate seasonal fluctuations in disease frequency. A more precise and less biased method for estimating seasonal factors consists of estimating the overall average disease incidence for the entire dataset and dividing that number into the mean disease frequency for each month (Zhang et al., 2014), e.g.,

$$\text{Seasonal index} = \frac{\text{Mean disease frequency for each month}}{\text{Averaged disease frequency}}$$

If the resulting seasonal index is greater than 1, it means that the incidence is usually higher than the average level. If it is lower than 1, it means that the incidence is usually lower than the average level. Once the seasonal indices are calculated, one can deseasonalize the raw time series data by dividing by the corresponding index, e.g.,

$$\text{Deseasonalized data} = \frac{\text{Raw data}}{\text{Seasonal index}}$$

The **long-term (secular) trend** is estimated from the deseasonalized data. There are many ways to estimate the long-term trend, such as moving average, exponential smoothing, and linear regression. **Twelve-month moving averages** can be used to smooth out or eliminate irregular variations and those with periodicities of 12 months or less. The result is an approximate secular trend line. **Least-squares regression** is a statistical technique that derives a line with the least mean squared deviation from all data points. Details and assumptions of the procedure can be found in standard statistical texts, and it is a standard option in spreadsheets and statistical software.

Deviations from the deseasonalized trend line comprise the combined cyclical and irregular variation (**residuals**). **Cyclical changes** refer to the rise and fall of disease frequency developing at intervals longer than 1 year. **Irregular variation** reflects random or unpredictable variation in disease occurrence among individuals in a population. Cyclical and irregular variation may be associated with disease outbreaks.

EXAMPLE 11.3: HOW DOES THE SEASONALITY OF CANINE LEPTOSPIROSIS COMPARE ACROSS THE UNITED STATES?

Background: Leptospirosis is a common zoonotic disease with worldwide distribution affecting many mammalian species. Dogs usually become infected by contact with urine or water containing *Leptospira*. Previous studies have reported a seasonal increased risk for leptospirosis, but there is no consistent seasonality reported across regions in the United States.

Objectives: Lee et al. (2014) evaluated and compared seasonal patterns in seropositivity for leptospirosis in dogs for four U.S. regions: northeast (NE), midwest (MW), south-central (SC), and California-southern west coast (CS).

Study Design: Cross-sectional.

Methods: Test results for a total of 44,916 canine serum samples submitted by practitioners to a commercial laboratory for microscopic agglutination tests (MAT) from January 1, 2000, through December 31, 2010, were included in the study. Positive cases were defined as MAT titers $\geq$1:3200 for at least one of seven *Leptospira* serovars. MAT results from four geographic regions were represented (Figure 11.3) and were included in regional analyses based on hospital zip code.

A seasonal-trend decomposition procedure for time series was used for the time series analysis, which decomposes the time series into three components: seasonal, trend, and remainder (residuals). Annual (12-month) patterns as the seasonal component were first determined, then removed for smoothing to find the trend. The remainder component was the monthly residuals from the seasonal plus trend fit.

Monthly variation in the seropositive rate was separately evaluated using a seasonal cycle subseries plot that assumed yearly periodicity. To statistically compare average monthly seropositive rates, univariate logistic regression models were constructed for each region with odds ratios and 95% confidence intervals calculated for each month. The lowest average monthly rate was used as the reference month.

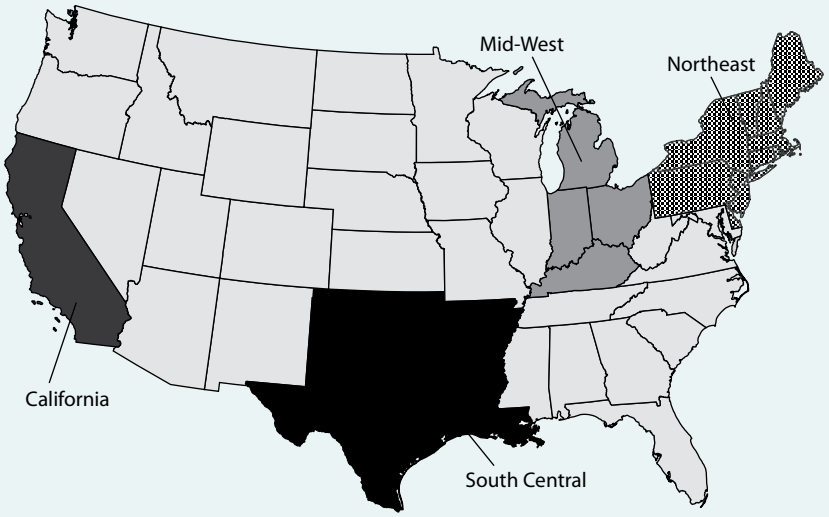

FIGURE 11.3 Four geographic regions used in assessment of temporal patterns of seropositivity to *Leptospira* in dogs, 2000–2010. (From Lee HS et al. *J Vet Intern Med* 2014;28:779–788. With permission.)

Results: A total of 2012 of 44,916 (4.48%) samples were seropositive. The time series plots of trend, seasonality, and remainder components indicated seasonality to be the strongest component in the overall data during the study period (Figures 11.4 and 11.5).

The seasonal cycle subseries plot showed that mean monthly seropositive rates were greatest in November, followed by December and October (Figure 11.6). Overall, rates were greatest in the fall (September–December), although a smaller peak in spring (May) was also noted. The seasonal cycle subseries plot exhibited interannual variation by month, as noted by the vertical bars. A large increase above average occurred in May of one year (2007) compared to other years.

Compared to seropositive rates for February, significantly higher monthly rates occurred during the second half of the year in the MW (OR 3.92–6.35) and NE (OR 2.03–4.80) regions,

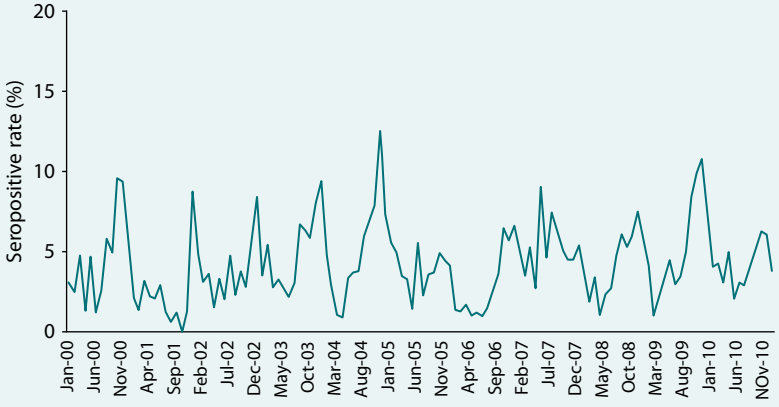

FIGURE 11.4 Seropositive rate (%) of microscopic agglutination tests using a 1:3200 titer cutoff for canine leptospirosis in the United States by month from January 2000 through December 2010. (From Lee HS et al. *J Vet Intern Med* 2014;28:779–788. With permission.)

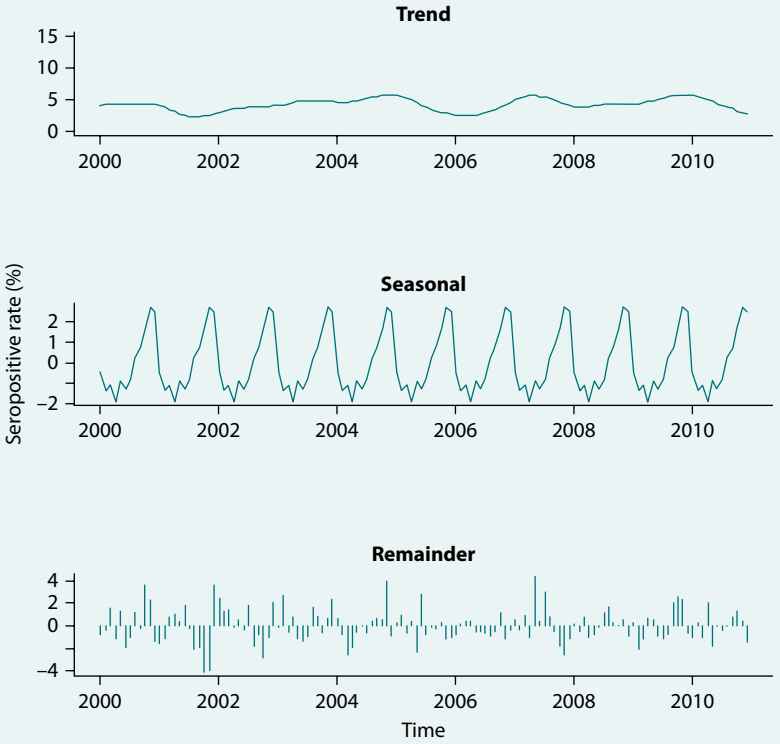

FIGURE 11.5 Seasonal-trend decomposition of the monthly seropositive rate (%) for canine leptospirosis in the United States, 2000–2010, displayed in its three components of trend, seasonal, and the remainder. (From Lee HS et al. *J Vet Intern Med* 2014;28:779–788. With permission.)

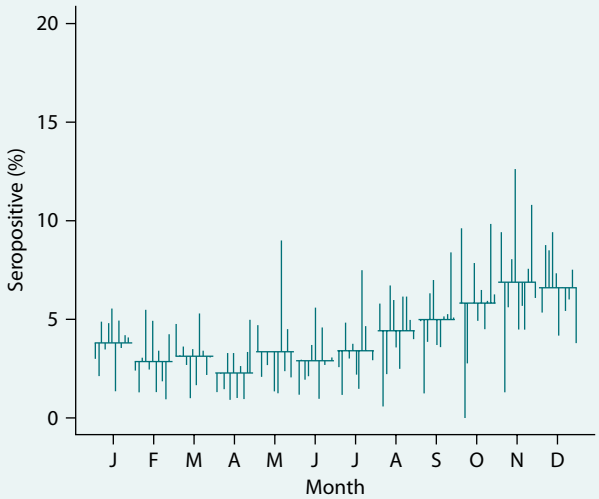

FIGURE 11.6 Seasonal cycle subseries plot of the monthly seropositive rate (%) for canine leptospirosis in the United States, 2000–2010. Horizontal lines display the overall mean seropositive rates for each month in the 11-year period. Each vertical line above or below the horizontal bar reveals the difference from the overall monthly average in each year of the data. (From Lee HS et al. *J Vet Intern Med* 2014;28:779–788. With permission.)

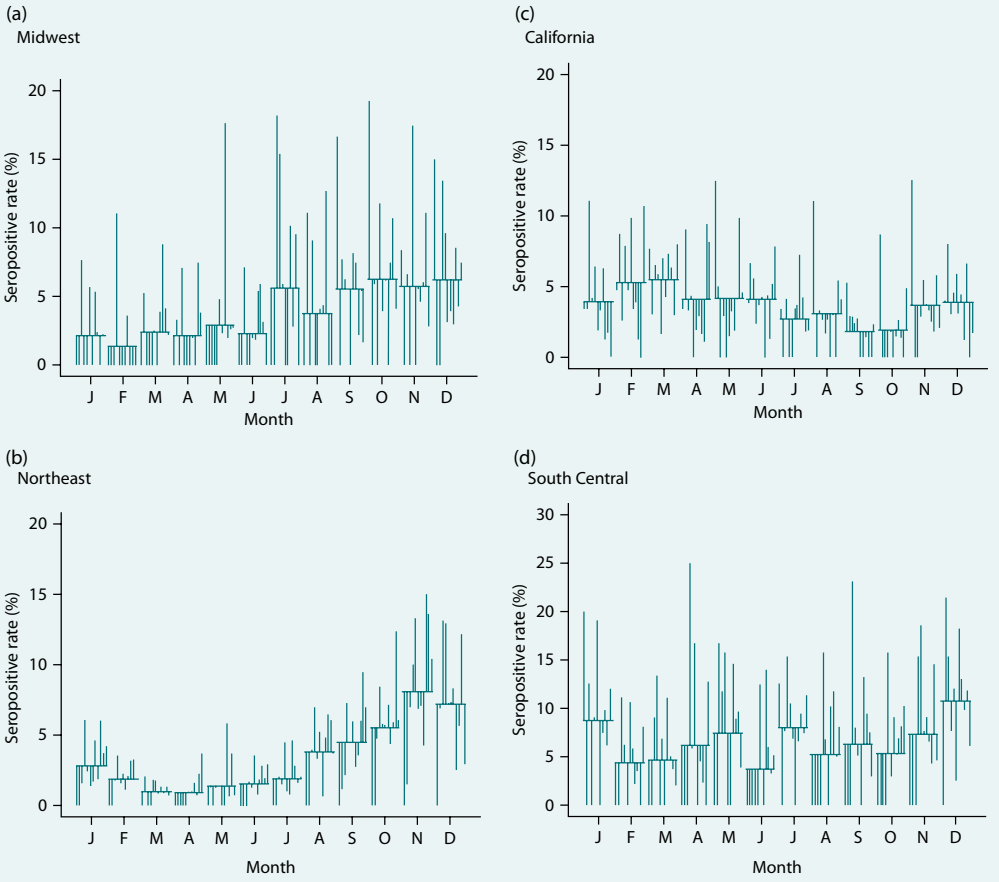

FIGURE 11.7 Seasonal cycle subseries plots of the seropositive rate (%) on a monthly basis for canine leptospirosis in four U.S. regions. (From Lee HS et al. *J Vet Intern Med* 2014;28:779–788. With permission.)

and only in January (OR 2.34) and December (OR 1.74) in the SC region. Monthly seropositive rates indicative of seasonality were observed earlier in the calendar year for both the CS and SC regions (Figure 11.7).

Conclusions and Significance: The authors concluded that seasonal patterns for seropositivity to *Leptospira* differed by geographic region. Although risk of infection in dogs can occur year-round, knowledge of seasonal trends can assist veterinarians in formulating differential diagnoses and evaluation of exposure risk.

FOLLOW-UP QUESTION 11.3

What aspects of the study design might confound the interpretation of the results? See Answer 11.3 at the end of this chapter.

11.3.4 Spatial Distribution

There are a number of ways to depict the spatial distribution of disease frequency. **Areal maps** depict the distribution and frequency of disease within defined areas or boundaries, as counties, states, or ecological zones. An example can be found in Figure 11.3. Another approach is the simple **spot (or**

dot) map, where each dot either represents a case or is scaled to represent the frequency of disease. There are many variations of spot maps, however, and one should always examine them carefully so as not to misinterpret the information provided. **Overlay mapping**, where two or more spatial distribution maps are superimposed on one another, provides a simple technique for exploring the association among spatially distributed variables.

REFERENCES

Cowled B, Ward MP, Hamilton S et al. The equine influenza epidemic in Australia: Spatial and temporal descriptive analyses of a large propagating epidemic. *Prev Vet Med* 2009;92:60–70.

Dicker RC, Coronado F, Koo D et al. *Principles of Epidemiology in Public Health Practice*. 3rd ed. Atlanta: Centers for Disease Control and Prevention (CDC); 2012. PDF version available at: https://www.cdc.gov/ophss/csels/dsepd/ss1978/SS1978.pdf. Outbreak investigation module available online at: https://www.cdc.gov/ophss/csels/dsepd/ss1978/lesson6/section2.html

Hartnack S, Doherr MG, Toutounian-Mashad AT et al. *Campylobacter* monitoring in German broiler flocks: An explorative time series analysis. *Zoonoses Public Health* 2009;56:117–128.

Lee HS, Levine M, Guptill-Yoran C et al. Regional and temporal variations of *Leptospira* seropositivity in dogs in the United States, 2000–2010. *J Vet Intern Med* 2014;28:779–788. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4895461/

Miller, MA, Conrad, PA, Harris M et al. A protozoal-associated epizootic impacting marine wildlife: Mass-mortality of southern sea otters (*Enhydra lutris nereis*) due to *Sarcocystis neurona* infection. *Vet Parasitol* 2010;172:183–194. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4907795/

Molina LL, Angón E, García A et al. Time series analysis of bovine venereal diseases in La Pampa, Argentina. *PLOS ONE* 2018;13:e0201739. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6078287/

Schwabe CW, Riemann HP, Franti CE, *Epidemiology in Veterinary Practice*. Philadelphia: Lea and Febiger; 1977.

Zhang X, Zhang T, Young AA et al. Applications and comparisons of four time series models in epidemiological surveillance data. *PLOS ONE* 2014;9:e88075. Full text available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0088075

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 11.1: This outbreak appears to have resulted from a complex interaction among wildlife, their parasites, and weather-related phenomena, none of which are easily mitigated. Further, as a relatively infrequent and unpredictable event, it would be difficult to justify significant investments of money and resources for future control efforts. Perhaps the best approach is to maintain regular surveillance of southern sea otter strandings to better understand the magnitude of the problem. Just such a system, the California Sea Otter Stranding Network (https://www.usgs.gov/centers/werc/science/california-sea-otter-stranding-network?qt-science_center_objects=0#qt-science_center_objects), has been operational since 1968.

Answer 11.2: Assuming that immunity would begin to develop 2 weeks following the first vaccination, the earliest date that vaccine-induced immunity may have played a role in limiting EI spread would have been October 13 to November 29, 2007. By this time, the epidemic would have been in decline.

Answer 11.3: The authors reported differences in patterns of seroprevalence between geographic regions. However, the regions used to group results are extensive and additional variations could exist within regions. There is a 7–10-day delay between exposure to *Leptospira* infection and development of detectable serum antibody. Thus, the month that an infection was actually acquired could have been earlier than the month that a positive test was reported. The reason for serologic testing was not determined. It was assumed that the request for a serologic test was predicated on clinical signs potentially associated with leptospirosis in dogs, rather than a request for titer testing after

vaccination. The relationship between seasonal seroprevalence and weather patterns in the regions studied is hypothetical, as actual meteorological data were not recorded.

The above observations were put forth by the authors. They do not invalidate the findings nor diminish the clinical relevance of this study, but they do suggest refinements for future follow-up studies.

12 Establishing Cause

12.1 INTRODUCTION

Epidemiologic investigation of a disease outbreak of unknown etiology will usually incriminate a number of factors, or **determinants**, of the disease. Usually only one factor (the etiologic agent) is causal, and its relationship to the disease syndrome may be confirmed by some variation of **Koch's postulates**. Other factors, termed host and environmental determinants, may facilitate the introduction and spread of the etiologic agent within animal populations. In this chapter, we examine how these determinants are identified and how their relationship to disease is established.

12.2 MULTIPLE CAUSATION OF DISEASE

Determinants of disease include both the etiologic agent(s) directly responsible for disease and other factors that facilitate exposure, multiplication, and spread in the population. These determinants can be categorized as **agent**, **host**, and **environment** (or management) factors. The way in which these factors interact to cause disease has been referred to as the **web of causation**, which is another expression of the concept of multiple causality.

> ***Determinants of disease include the agent, host, and environment.***

12.2.1 Agent Factors

The biological properties of agents, such as pathogenicity and virulence, strains, and genetic variability, are primary determinants of the ability of an agent to cause disease. **Pathogenicity** refers to the ability of an organism to cause disease (i.e., harm the host), whereas **virulence** refers to the degree of pathology caused by the organism. Factors contributing to the pathogenicity and virulence of disease agents are numerous and best explored with reference to specific organisms.

12.2.2 Host Factors: Susceptibility

The susceptibility of individual animals to disease is a second determinant of disease occurrence. Natural variation affects the response of individual animals to exposure to a disease agent. Most of the statistical examples that were discussed earlier have focused on this type of variation. Some animals have **innate resistance** to infection or disease due to age, sex, or breed. **Acquired resistance** in the individual may be the result of prior natural or artificial (vaccination) exposure to the agent. In some cases, animals are latently infected with an agent that has the potential to cause clinical disease. The triggering mechanism may be an altered immune response brought on by stress. An example is the predictable outbreak of "shipping fever complex" seen in cattle and horses shortly after being transported to a new location.

Populations also differ in susceptibility. Resistance in populations is called **herd immunity** (or **population immunity**) and is related to the proportion of resistant animals in the population. **Innate herd immunity** reflects a population that is resistant to an infection for some reason other than prior natural exposure or immunization. **Acquired herd immunity** results from the development of protective immunity in a population after natural exposure or immunization.

Populations differ in susceptibility. Resistance in populations is called herd immunity: and is related to the proportion of resistant animals in the population.

Increased herd immunity has the effect of limiting the spread of directly transmitted diseases by reducing the proportion of **effective contacts**, i.e., contacts between infected and susceptible animals that result in transmission of a disease agent (Smith, 2010; Fine et al., 2011). Increased herd immunity may also limit the spread of indirectly transmitted and airborne disease agents by reducing environmental contamination.

It follows that the higher the **basic reproduction number** ($\mathbf{R_o}$) of a pathogen, defined as the number of secondary cases that one case would produce *in a completely susceptible population* (Delamater et al., 2019), the higher the level of herd immunity that must be achieved for its eradication. If the exposed population is not completely susceptible, either through innate or acquired immunity, then the transmissibility of the disease agent is reduced and expressed as the **effective reproduction number** ($\mathbf{R_e}$, or simply **R**) (Smith, 2010). In this case, the value of R can be estimated using the equation $R = (1 - P) \times R_o$, where (P) is the proportion of the population that is immune to infection, and $(1 - P)$ the proportion of susceptibles. If R can be reduced through vaccination below that required for maintenance of a disease agent in the population ($R \leq 1$), its eventual eradication might be achieved. A rearrangement of this equation, $(R_o - 1)/R_o$, can be used to directly estimate the

TABLE 12.1

Relationship between a Pathogen's Basic Reproduction Number (R_o) and the Proportion of the Host Population that Must Be Vaccinated (Herd Immunity) to Achieve Eradication of Some Directly and Indirectly Transmitted Human Diseases

Disease	Location and Time of Data Collection	R_o	P (%)
Smallpox	Developing countries before global eradication campaign	3.5	72
Measles	England and Wales (1950–68)	16–18	95
	Ghana (1960–68)	14–15	94
Pertussis	England and Wales (1944–78)	16–18	95
	Maryland (1908–17)	13	93
German measles	England and Wales (1979)	6	84
	West Germany (1972)		
Chicken pox	Parts of United States (1913–21; 1943)	7–11	91
	England and Wales (1944–68)	10–12	92
Diphtheria	Parts of United States (1910–19)	4–5	80
Scarlet fever	Parts of United States (1908–19)	6–7	86
Mumps	England and Wales (1960–80)	11–14	93
	Netherlands (1970–80)		
Poliomyelitis	United States (1955)	5–6	84
	Netherlands (1960)	6–7	86
Malaria (*Plasmodium falciparum*)	Northern Nigeria (1970s)	80	99
Malaria (*Plasmodium malariae*)	Northern Nigeria (1970s)	16	94
HIV (Type I)	England and Wales (male homosexuals; 1981–85)	2–5	80
	Kampala, Uganda (heterosexuals; 1985–87)	10–11	91

Source: May RM. *Am Scientist* 1983;71:36–45. With permission.

R_o = the number of secondary infections produced by one case in a totally susceptible population.

P (%) = the proportion of the population that must be protected by immunization to achieve eradication, i.e., $R_o (1 - P) \leq 1$.

proportion of the population that must be protected to achieve eradication. This proportion has been referred to as the "**herd immunity threshold**" (Fine et al., 2011).

Very high levels of artificially induced herd immunity are required to eradicate diseases whose basic reproduction numbers are high. Table 12.1 (May, 1983) summarizes historical data on the relationship between estimates of R_o for several pathogens and the level of herd immunity required to eradicate the disease. Although some of these values for R_o have been adjusted in light of more recent findings (Smith, 2010), they serve to illustrate the relationship between R_o and the level of herd immunity required to achieve eradication. The relatively small value of R_o (3–5) for smallpox and corresponding low level of herd immunity that must be artificially induced may partially explain the success of the global eradication campaign. Other factors are the obviousness of the disease and availability of an effective vaccine. In contrast, the high values of R_o for *Plasmodium falciparum* malaria (∼80) suggest that eradication through vaccination will be much more difficult to achieve. Furthermore, carriers may easily escape detection, and prototype vaccines do not prevent infection, only disease. The cyclical nature of disease epidemics in domestic animals and wildlife may be related to the destabilizing effect of population fluctuations upon herd immunity.

EXAMPLE 12.1: WHAT IS THE CRITICAL VACCINATION THRESHOLD FOR CONTROLLING RABIES IN FREE-ROAMING DOGS?

Background: Rabies is a serious yet neglected public health threat in resource-limited communities in Africa, where the virus is maintained in populations of owned, free-roaming domestic dogs. Rabies elimination can be achieved through the mass vaccination of dogs, but maintaining the critical threshold of vaccination coverage for herd immunity (herd immunity threshold) in these populations is hampered by their rapid turnover. Knowledge of the population dynamics of free-roaming dog populations can inform effective planning and implementation of mass dog vaccination campaigns to control rabies. The basic reproduction number (R_o) for canine rabies worldwide has been estimated as $\leq$1.72, yielding a critical herd immunity threshold of ≈40% (Hampson et al., 2009). Thus, theory and empirical evidence predict that outbreaks of rabies in dogs can be controlled if at least 40% of the population is immune at any time. The World Health Organization recommends that to achieve control and eventual elimination of dog rabies, programs must ensure that mass dog vaccination campaigns achieve a vaccination coverage of at least 70% of the population in a given area, and that such campaigns must recur usually annually. Following vaccination campaigns, the proportion of immune individuals in the population declines as vaccinated dogs die and susceptible dogs enter the population through birth or migration. The actual target vaccination coverage to be achieved during campaigns is thus dependent on the demographic dynamics of the canine population, as well as the interval between campaigns and duration of vaccine-induced immunity. A number of studies of free-roaming dog populations in communities in Africa have revealed that, despite appearances, there is little evidence for the presence of large numbers of unowned dogs in these populations.

Objectives: Conan et al. (2015) designed a study to quantify demographic parameters covering all owned dogs in a rabies-affected, resource-limited community in South Africa, particularly the core demographic rates of births, deaths, and migrations, and to assess the implications of dog population dynamics for rabies control through mass vaccination.

Study Design: Cohort study.

Methods: A canine health and demographic surveillance system (HDSS) was implemented to monitor the entire owned dog population within a defined geographic area (demographic surveillance area; DSA) in a community in Mpumalanga Province, South Africa. Demographic

rates were quantified over a 24-month period, from January 1, 2012, through January 1, 2014, and their implications for rabies control assessed by simulating the decline in vaccination coverage over time. To uniquely and permanently identify individual dogs, microchips were subcutaneously implanted into dogs present at the start of the study and into those dogs that entered the population during the study period. Dogs that could not be handled to implant a microchip were assigned a unique identification code. Dogs were also identified by name and appearance. All dogs enrolled in the study were photographed. Following the census (Round 1), five follow-up rounds (Rounds 2–6) were conducted from December 2011 through May 2014, resulting in all households being visited approximately every 6 months during this period. All households in the DSA were visited during each round. All owned dogs were recorded at each visit, as were the demographic events that occurred in the period between visits, including births, deaths, and migrations into and out of households. Data including age, sex, and rabies vaccination status were collected during the census for all dogs and at each round for new dogs. Vaccination history was based on owners' reports.

Results: During this period, there were substantial fluctuations in the number of owned dogs and crude canine birth and death rates. The dog population declined by 10%. Annual population growth rates were +18.6% in 2012 and −24.5% in 2013. Crude annual birth rates (per 1000 dog-years of observation) were 451 in 2012 and 313 in 2013. Crude annual death rates were 406 in 2012 and 568 in 2013. Females suffered a significantly higher mortality rate in 2013 than males (mortality rate ratio; MRR = 1.54, 95% CI = 1.28–1.85). In the age class 0–3 months, the mortality rate of dogs vaccinated against rabies was significantly lower than that of unvaccinated dogs (2012: MRR = 0.11, 95% CI = 0.05–0.21; 2013: MRR = 0.31, 95% CI = 0.11–0.69).

The net migration rate, measured as the total number of in-migrations to households minus the total number of out-migrations from households, was 12.3% in 2012 and −2.1% in 2013. Most dogs entered households through birth (61%) or as gifts (31%) and exited through death (71%) or being given away (21%).

To assess whether a 70% vaccination coverage achieved during annual campaigns is sufficient to maintain population immunity above the critical threshold of 40% for a 12-month period, the authors evaluated owner-reported vaccination coverage and two simulated vaccination campaigns, one on January 1, 2012, and another on January 1, 2013. For the simulation, a positive vaccination status was randomly assigned to 70% of all dogs present in the population on those dates. The results of the simulation showed that achieving a 70% vaccination coverage during annual campaigns would maintain coverage above the critical threshold of 40% for at least 12 months. The actual minimum vaccination coverage needed for the campaigns was 61% in 2012 and 52% in 2013, based on measured demographics and an assumed duration of post-vaccinal protection of 3 years. Vaccination coverage based on owner-reported vaccination history was 33% before the start of the house-to-house vaccination campaign in 2012 and 78% following completion of the first round of the campaign, well above the threshold level of 40% threshold until a second vaccination campaign in 2013.

Conclusions and Significance: This study provides an evidence base for the World Health Organization's empirically derived target of 70% vaccination coverage during annual canine rabies vaccination campaigns. Vaccinating 70% of the population during annual campaigns would be sufficient to maintain herd immunity to rabies in the period between campaigns. Achieving this will be effective even in highly dynamic populations with extremely high growth rates and rapid turnover. This increases confidence in the feasibility of dog rabies elimination in Africa through mass vaccination.

FOLLOW-UP QUESTION 12.1

The authors conclude that yearly mass dog vaccination campaigns that reach 70% of the population will be effective in bringing rabies under control and can contribute to rabies elimination, even in canine populations undergoing extremely high growth rates and rapid turnover. Can you think of an information need that is critical to the success of such a program? See Answer 12.1 at the end of this chapter.

12.2.3 Environmental (Management) Factors

According to most general practitioners, environmental or management factors are the most important determinants of disease occurrence. Management factors also comprise a category of factors that are difficult to quantify and manipulate. Examples are the influence of milking hygiene on the occurrence of bovine mastitis or management practices on neonatal calf mortality.

12.3 SOURCES OF BIAS IN EVALUATING CAUSE-EFFECT RELATIONSHIPS

Conducting an epidemiologic investigation is much like playing the role of Inspector Hercule Poirot in an Agatha Christie mystery. Observational data (evidence at the crime scene) is used retrospectively (after the fact) to identify a causal association (the guilty) and infer the nature of that association (the motive). The problem for both the detective and the epidemiologist is that the investigation may implicate several causal variables that are directly or indirectly associated with the outcome. Ignoring the potential relationships among variables can lead to biased estimates of the effect of exposure on outcome. Investigators have several tools at their disposal for sorting out these effects, which are discussed below.

Conducting an epidemiologic investigation is much like a criminal investigation. Several causal variables that are directly or indirectly associated with the outcome may be implicated. Ignoring the potential relationships among these variables can lead to biased estimates of the effect of exposure on outcome.

12.3.1 Confounding

Confounding occurs when two or more variables found to be associated with an outcome (as disease) are also associated with each other. As a result, it is impossible, from a cursory examination, to determine which variable is responsible for the observed outcome. The confounder's association with the outcome may be causal, or it may simply be associated with the true cause. For example, prior to the discovery of the role of freshwater stream snails and their associated trematode parasites and aquatic insects as intermediate hosts and vectors for *Neorickettsia (Ehrlichia) risticii*, the causative agent of equine monocytic ehrlichiosis (Potomac horse fever) (Baird and Arroyo, 2013), a number of other factors were believed to contribute in some way to the risk of contracting the disease. These included ticks, biting flies, mosquitoes, white-footed mice, and duration of access to pasture. It is now apparent that the relationship of these previously suspected factors was only through a passive association with the true cause.

12.3.2 Interaction or Effect Modification

Interaction occurs when one variable modifies the effect of another. For example, if the severity of disease from exposure to an agent is greater in the very young or very old, we say that there is an

interaction between the two variables, or **effect modification** of age upon disease. Other common variables that can modify the magnitude of cause-effect relationships are breed and sex. Interaction can be detected by performing stratified (**Mantel-Haenszel**) two-by-two analyses, stratifying on the suspected effect modifier. For example, if we wish to evaluate the effect of a particular agent on calf survival from birth to weaning, we might construct a series of two-by-two tables stratified by age groups. If there is a notable difference in risk (relative risk or odds ratio) between one or more age groups (**strata**), then interaction between age and pathogenicity should be considered.

If interaction among predictor variables exists, then the Mantel-Haenszel summary odds ratio, which assumes uniform risk among subgroups or strata, may be biased. If more than one major predictor variable (risk factor, exposure) exists, or if interaction is present, then multivariate logistic regression (see Chapter 6) should be used to estimate the contribution of each risk factor to the outcome.

12.3.3 Multicollinearity

Multicollinearity occurs when two or more predictor variables (**covariates**) are also highly correlated with each other, independent of the outcome variable. Examples might be the use of erythrocyte counts and hemoglobin levels to assess anemia, or two related serologic tests as predictors of the same disease. In each case, the two predictor variables are really measuring the same thing, such that we would expect values for both to increase or decrease together. Many medical measurements are highly correlated, and it is standard practice to perform several "corroborating" tests in diagnosing an illness.

Multivariate logistic regression, which is often used to estimate the contribution of multiple risk factors to disease outcomes, cannot distinguish between two predictor variables that are highly correlated. For this reason, univariate (two-by-two) analysis should always be performed first to identify those predictor variables associated with the outcome. The relationship between individual predictor variables can then be assessed with two-by-two tables or by examining a correlation matrix of predictor variables. The presence and degree of multicollinearity can be assessed by asking how well each independent (X) variable can be predicted from the other X variables (ignoring the Y variable).

12.3.4 Procedure for Evaluating Interaction and Confounding

For the following discussion, it is assumed that the odds ratio is the parameter of interest (see Chapter 6). This is usually the case when evaluating the contribution of a number of potential risk factors to disease occurrence.

Step 1: Test the crude association of the explanatory variable with the outcome or response variable, i.e., conduct a chi-square test to evaluate significance of the association and calculate a crude odds ratio (along with its confidence interval) to measures the strength of the association. This step provides a baseline reference statistic for comparison with the results of subsequent stratified analyses.

Step 2: The simplest analytic tool for evaluating interaction and confounding is the Mantel-Haenszel stratified analysis. It divides the analysis into separate tables for each value of the "stratifying" variable (such as age group), and then combines the results in a way (referred to as the **adjusted odds ratio**) that removes the effect of confounding. By stratifying on a variable, we eliminate the effect of confounding by that variable. The adjusted OR is only valid, however, if the ORs in individual strata are similar in value. **If they are not, then interaction is present.** If interaction is present, then two options are available:

1. Set up the stratification in a different way by combining groups, etc. to reduce the number of differing strata to a minimum.
2. If strata still differ, report the result for each of the differing strata (such as age groups) separately, rather than in a combined result.

Step 3: If there is no statistically significant interaction among predictor variables, then the next question is whether the stratifying variable confounds the exposure-disease relationship. To assess confounding, the crude OR (estimated with the chi-square test in Step 1 above) is compared with the Mantel-Haenszel adjusted OR. The adjusted OR is a weighted measure of association and is always a more valid estimate of the "true" OR than the crude OR. There is no statistical test for confounding. Some investigators may choose an arbitrary rule for deciding whether the level of confounding is "important" by comparing the crude and adjusted parameters. For example, the decision may be: if the crude and adjusted parameters differ by more than 5% or 10%, the stratifying variable will be considered a confounder. If the stratifying variable does not modify or confound an exposure-disease relationship, then it can be ignored in any further analyses.

In summary, if the odds ratios for strata in a series of stratified tables are not similar, then interaction between the stratifying factor and the risk factor is present. In this case, one could set up the strata differently or just present the stratum-specific estimates separately. Logistic regression or other multivariate methods may be helpful unless the number of subjects is too small to draw definite conclusions. Regardless, the data should be thoroughly explored using stratified analysis before undertaking a logistic regression analysis. If the number of confounding factors is fairly small and the odds ratios are homogeneous from stratum to stratum, stratified analysis may be all you need (CDC, 2016).

Several online tools are available for calculating crude and adjusted odds ratios. *Statulator* is a free on-line statistical calculator that is extremely simple and user friendly. The *Statulator Chi-Square Test* (http://statulator.com/stat/chisq.html) calculates the chi-square test for a 2×2 table, the crude odds ratio, and relative risk with confidence intervals, and interprets the results. *Statulator Stratified Analyses* (http://statulator.com/stat/StratifiedAnalysis.html) calculates the stratified odds ratio or relative risk and tests their homogeneity, conducts a Mantel-Haenszel pooled odds ratio or relative risk and evaluates its significance, and interprets the results. The reader is encouraged to apply these tools to actual or hypothetical data to better appreciate insights gained from these analyses.

12.3.5 The Choice of Multivariate versus Stratified Analysis

When should multivariate logistic regression be used? Answer: if the outcome variable is dichotomous (e.g., ill or not ill) and there is more than one major predictor variable (risk factor, exposure) of interest, or if interaction is present. With only one predictor variable and no interaction, Mantel-Haenszel stratified analysis can compensate for the confounding if the number of confounders is small. If there are many confounders, the number of strata becomes large and each one contains small numbers, so that Mantel-Haenszel analysis becomes impractical.

As a general rule, simple and stratified univariate analyses should be done before embarking on logistic regression. It gives a feeling for the dataset in a simple setting.

12.4 ESTABLISHING CAUSE

In 1882, Koch set forth the following postulates for determining that an infectious agent is the cause of a disease (Fletcher et al., 2014):

1. The organism must be present in every case of the disease.
2. The organism must be isolated and grown in pure culture.
3. The organism must cause the specific disease when inoculated into an animal.
4. The organism must then be recovered from the animal and identified.

Koch's postulates were an important step in removing disease causation from the anecdotal evidence and superstitions of the time. However, the causes of many diseases cannot be established

by means of Koch's postulates. In light of this, proposals have been put forth for an alternative set of "molecular Koch's postulates," which could be applied to microbial studies designed to examine the role of specific genes and their products in the pathogenesis of infection and disease (Breitschwerdt et al., 2013). There has also been a proposal to divide pathogenic infectious agents into two major categories: "frontal pathogens" and "stealth pathogens" (Merrell and Falkow, 2004). Frontal pathogens typically induce an acute infection, while stealth pathogens are characterized by persistent infections over protracted periods of time. Examples of stealth pathogens are *Helicobacter pylori*, a cause of ulcers in humans, and *Bartonella henselae*, the cause of cat scratch disease. Causation associated with frontal pathogens can in most instances be established using the original Koch's postulates. In contrast, establishing causation for stealth pathogens, particularly in the context of chronic disease causation, is more likely to be successful when the more recently proposed molecular Koch's postulates are applied to investigations of disease pathogenesis.

The causes of many diseases cannot be established by means of Koch's postulates.

EXAMPLE 12.2: WHAT FACTORS CONTRIBUTE TO THE INCIDENCE OF SHIPPING FEVER IN HORSES DURING INTERNATIONAL AIR TRANSPORT?

Background: Shipping fever (SF) is a common respiratory disease of horses associated with transport over long distances by road, rail, sea, or air. When SF is identified and treated in the early stages of the infection, affected horses usually recover quickly. However, if the infection spreads and pleuropneumonia develops, intensive medical treatment is necessary and the prognosis for a horse returning to its former level of activity becomes poor. Horses with severe SF may develop secondary complications, such as pulmonary abscessation, colitis, and laminitis, which can be fatal.

Shipments of horses arrive continuously in Hong Kong (HK) throughout the year, primarily for thoroughbred flat racing, with flights originating from Australia, New Zealand (NZ), the United Kingdom (UK), or the United States of America (USA). Due to the long flight times associated with travel from these destinations, horses imported into HK are at a risk of developing SF. Few studies have been performed to identify risk factors for SF that could be specific to air transport and guide preventative measures.

Objectives: Hurley et al. (2016) conducted a 2-year prospective study between February 2011 and January 2013 to determine the incidence and risk factors for shipping fever (SF) in horses transported by air to HK.

Study Design: Cohort study.

Methods: The study analyzed data from 869 horses on 81 flights arriving from Australia (n = 24), New Zealand (n = 18), the United Kingdom (n = 33), and the United States of America (n = 6), with a median of 10 horses per shipment (range 1–48). Using a questionnaire, data were collected from professional flying grooms regarding the journey to HK and horses in the shipment. Horses were monitored in quarantine for 2 weeks after arrival in HK and clinical signs of SF recorded. Horses were diagnosed with SF if they developed one or more clinical signs of lower respiratory tract infection, including fever (>38.6°C), inappetence, depression, coughing, nasal discharge, increased respiratory rate or effort, and adventitious lung sounds. Outcome variables were assessed at the horse and shipment levels. The horse-level outcome variable was the presence of SF in an individual horse. The shipment-level outcome variable

TABLE 12.2
Multivariable Poisson Regression Results for the Outcome of the Rate per Shipment of Shipping Fever in Horses Traveling to Hong Kong

Variable	Level	Incidence Rate Ratio	95% CI	*p*
Country	Australia	REF		
	United Kingdom	3.08[a]	1.60–5.93	<0.001
	New Zealand	2.40	1.22–4.71	0.01
	United States of America	2.43	0.66–8.89	0.18
Month	January	REF		
	February	2.27	0.71–7.28	0.17
	March	5.61	1.55–20.31	0.01
	April	1.65	0.40–6.78	0.49
	May	4.51	1.43–14.26	0.01
	June	0.47	0.05–4.23	0.50
	July	2.65	0.85–8.29	0.09
	August	2.75	0.90–8.34	0.08
	September	2.52	0.80–7.89	0.11
	October	2.18	0.69–6.88	0.19
	November	1.91	0.55–6.69	0.31
	December	0.65	0.07–5.94	0.70

Source: Hurley MJ et al. *Vet J* 2016;214:34–39. With permission.
Abbreviation: 95% CI, 95% confidence interval.
[a] *Interpretation*: The rate per shipment of SF in shipments from the United Kingdom is 3.08 (95% CI 1.60–5.93) times that when shipments arrive from Australia.

was the number of horses with SF per consignment. Poisson and logistic regression models were used to identify risk factors for SF at the horse and shipment levels.

Results: During the 2-week quarantine period, 94/869 (10.8%) horses, from 49/81 (60%) flights, developed clinical signs of SF. All SF cases in the current study were mild and resolved without complication following treatment. The relative risks (incidence rate ratios) of SF in shipments from the United Kingdom (3.08) and New Zealand (2.40) were significantly higher ($p < 0.01$) than for shipments from Australia (Table 12.2). Shipments arriving in HK during March and May were 5.61 and 4.51 times more likely, respectively ($p < 0.01$), to include horses that developed SF compared to shipments arriving in January.

Conclusions and Significance: While factors such as country of origin cannot be altered, extra care of horses at a higher risk or with other risk factors for SF can be managed, with particular emphasis on the early detection of the condition to reduce the effect that SF has on the health and welfare of horses arriving into HK. The identification of these risk factors and the recognition of at-risk shipments will help focus attention on preventative strategies.

FOLLOW-UP QUESTION 12.2

What factors may contribute to the role of country of origin and season on the incidence of SF in transported horses? See Answer 12.2 at the end of this chapter.

As illustrated in the previous example, the "cause" is not a single etiologic agent but rather a triad of (1) management-related stress factors plus (2) a primary infection by any of several agents, followed by (3) a superinfection with any of a variety of additional agents. For most disease syndromes, there are many potential causes, and a single etiologic agent may cause a disease syndrome common to several other diseases. Koch's postulates are useful only in those special circumstances in which one particular cause dominates and when that cause is physically transmissible. The underlying assumption of Koch's postulates, one cause–one disease, is too simplified (Fletcher et al., 2014). Fortunately, other criteria may be applied to test the strength of a presumed cause-effect relationship. A description of these criteria follows.

12.4.1 Strength of Study Design

In Chapter 1, a variety of epidemiologic study designs were described. Generally, as one goes down the list in Table 1.5, the relative strength of study designs increases. Generally speaking, we can be more confident that a causal association exists as the strength of the study design increases.

12.4.2 Temporal Relationship between Cause and Effect

Demonstration of a **temporal relationship** between a hypothesized cause and effect is fundamental for concluding that a causal association exists. It is difficult to establish a temporal relationship in cross-sectional studies, in which both the outcome and suspected cause are measured at the same time. Longitudinal studies are particularly well suited for demonstrating causal associations, even if only two sampling periods occur. **Paired sampling** is a technique that has proven useful in establishing cause in clinical practice and outbreak investigation.

12.4.3 Strength of the Association

The stronger the association between a presumed causal factor and outcome, the more likely that a cause and effect relationship exists. As discussed in previous chapters, the **strength of association** between variables can be assessed by estimating relative risk, odds ratios, and correlation coefficients and through a number of statistical tests. However, we should not forget that association is a statistical concept that does not necessarily imply a cause-effect relationship. The case for causation can be strengthened if statistical associations also make biological sense (see "Biological plausibility" below).

12.4.4 Dose-Response Relationship

A cause-effect relationship is more likely to exist if it can be shown that varying amounts of the suspected cause are related to varying amounts of the effect. This is termed a **dose-response relationship**, or **biological gradient**. Dose can be measured in terms of absolute quantities, such as exposure to variable amounts of a substance or length of time over which exposure has occurred.

12.4.5 Biological Plausibility

Epidemiologic study designs are especially appropriate for the study of risk and prognostic factors (including treatment responses) for naturally occurring disease. Epidemiologic studies cannot, however, prove that a cause-effect relationship exists, only that an association exists that is unlikely to have arisen by chance alone. Statistical correlation does not prove causality. Research on mechanisms of disease provides the biological basis for believing that associations are, in fact, causal. On the other hand, information derived from research on mechanisms of disease cannot assume that a particular phenomenon will behave in nature as it does in the laboratory. For this, epidemiologic

studies must be conducted. Absence of a biological explanation does not necessarily mean that a causal association is absent. It may simply mean that current medical knowledge is incomplete.

12.4.6 Consistency

Evidence for a causal relationship is strengthened when several studies conducted under different conditions all come to the same conclusion. On the other hand, inconsistency in clinical findings may sometimes be attributed to differences in study design.

12.4.7 Elimination of Other Possibilities (Rule Out)

A differential list ranks the possible causes for an observed disease or other outcome. Sometimes the cause of disease, or a disease outbreak, is suggested by our inability to rule it out from a differential list of possible causes.

12.4.8 Reversible Associations

If removal of a factor results in decreased risk or frequency of disease, then it is more likely to be causal. This concept is the basis for current approaches to therapy and clinical trials.

EXAMPLE 12.3: HOW WELL DO EPIDEMIOLOGICAL CRITERIA EXPLAIN THE ROLE OF AIR TRANSPORT IN THE INCIDENCE OF SHIPPING FEVER IN HORSES?

Example 12.2 above presented the results of a study to identify potential causal (risk) factors for shipping fever in horses following air travel to Hong Kong. The authors applied six of the eight preceding criteria to assess whether associations between a number of explanatory variables and outcome (SF) were likely to represent a cause-and-effect relationship.

1. *Strength of study design*: The prospective study design permitted incidence rather than prevalence to be calculated, expanding the number and precision of statistical procedures that could be employed to express and analyze results.
2. *Temporal relationship between cause and effect*: The prospective study design also confirmed the likelihood that the outcomes occurred after air transport rather than having been already present at the time of shipment.
3. *Strength of the association:* The relationship between potential explanatory variables and outcome was assessed several ways and their statistical significance expressed as *p*-values.
4. *Biological plausibility:* The roles of stress, protracted immobility, seasonal weather variations in the countries of origin, fluctuations in relative humidity and ambient temperature during transport, and comingling with other horses during long flights with stopovers have been previously documented as plausible contributors to SF.
5. *Consistency*: Previous studies have documented the risk of SF in horses following air transport, but few studies have sought to identify risk factors for SF that could be specific to air transport.
6. *Dose-response relationship*: Travel time was explored as a surrogate for "dose," but was not confirmed as a statistically significant risk factor for SF.

FOLLOW-UP QUESTION 12.3

What explanatory variable could have been included to strengthen the association between air transport and risk of a horse developing SF? See Answer 12.3 at the end of this chapter.

REFERENCES

Baird JD and Arroyo LG. Historical aspects of Potomac horse fever in Ontario (1924–2010). *Can Vet J* 2013;54:565–572. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3659452/

Breitschwerdt EB, Linder KL, Day MJ et al. Koch's postulates and the pathogenesis of comparative infectious disease causation associated with Bartonella species. *J Comp Pathol* 2013;148:115–125.

Conan A, Akerele O, Simpson G et al. Population dynamics of owned, free-roaming dogs: Implications for rabies control. *PLoS Negl Trop Dis* 2015;9:e0004177. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4636342/

CDC. EpiInfo 7 User Guide, Version 3. Centers for Disease Control and Preventions, Atlanta, 2016. Available online at: https://www.cdc.gov/epiinfo/pdfs/UserGuide/EI7Full.pdf

Delamater PL, Street EJ, Leslie TF et al. Complexity of the basic reproduction number (R_0). *Emerg Infect Dis* 2019;25:1–4. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302597/

Fine P, Eames K, and Heymann DL. "Herd immunity": A rough guide. *Clin Infect Dis* 2011;52:911–916. Full text available at: https://academic.oup.com/cid/article/52/7/911/299077

Fletcher RH, Fletcher SW, and Fletcher GS. *Clinical Epidemiology—The Essentials*, 5th ed. Baltimore: Lippincott Williams & Wilkins, Baltimore; 2014.

Hampson K, Dushoff J, Cleaveland S et al. Transmission dynamics and prospects for the elimination of canine rabies. *PLoS Biol* 2009;7:e1000053. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2653555/

Hurley MJ, Riggs CM, Cogger N et al. The incidence and risk factors for shipping fever in horses transported by air to Hong Kong: Results from a 2-year prospective study. *Vet J* 2016;214:34–39.

May RM. Parasitic infections as regulators of animal populations: The dynamic relationship between parasites and their host populations offers clues to the etiology and control of infectious disease. *Am Scientist* 1983;71:36–45.

Merrell DS and Falkow S. Frontal and stealth attack strategies in microbial pathogenesis. *Nature* 2004;430:250–256.

Smith PG. Concepts of herd protection and immunity. *Proced Vaccin* 2010;2:134–139. Full text available at: https://www.sciencedirect.com/science/article/pii/S1877282X10000299?via%3Dihub

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 12.1: An accurate current estimate of the canine population is critical to assuring adequate vaccine coverage. The authors suggest that demographic surveillance of the entire owned, free-roaming dog population is feasible and can provide reliable, accurate data that are needed to determine the required level of coverage. The incorporation of a simple dog census/household survey into the house-to-house mass vaccination campaign, as practiced in the study area, could provide this information.

Answer 12.2: Besides country of origin and month of shipment, shipment-level explanatory variables include the number of horses in the shipment, the date of departure and arrival, the thermostat setting in the cargo hold, position of the horses in the aircraft, total time in the aircraft, and total travel time between quarantine units, as well as details of any delays and stopovers that occurred during the shipment. Horse-level explanatory variables include the age and sex of the horse and whether it was imported for racing or for use at one of the riding schools located in HK. Some of these variables approached statistical significance in the analysis, while others could not be assessed for lack of data.

Answer 12.3: Interestingly, the identification of potential causative agent(s) was not included among the explanatory variables for biological plausibility. If microbiological testing had been performed prior to shipment and during the post-arrival quarantine period, it would have strengthened the conclusion that observed signs were associated with the opportunistic replication of an infectious agent.

13 Source and Transmission of Disease Agents

13.1 SOURCES OF INFECTION

13.1.1 Iatrogenic Illnesses

A hospital-acquired infection, also known as a **nosocomial infection**, is an infection that is acquired in a hospital or other health care facility. Nosocomial infections with antimicrobial-resistant pathogens are of particular concern in the veterinary practice. **Iatrogenic illnesses**, i.e., those illnesses induced in a patient by a clinician's actions, extend the concept of nosocomial infections one step further by including any clinician-induced illness, infectious or otherwise, regardless of where it was acquired. Drug overdoses, the inappropriate use of particular therapeutic regimens, adverse drug reactions, and inadvertent surgical mishaps are examples of iatrogenic illnesses.

In some cases, as when attenuated vaccines are used, reactions are unavoidable. In these cases, the owner is advised that the patient may exhibit a brief period of mild illness following vaccination. Occasionally, however, a vaccine strain is suspected as the cause of an outbreak. Given the ubiquity of disease agents in the environment, it is often difficult to directly implicate a vaccine as the source of a disease agent. The recent availability of tools for the molecular characterization of microorganisms has given birth to a new branch of epidemiology—**molecular epidemiology**—that may be employed to trace the origin of a particular isolate.

Iatrogenic illnesses extend the concept of nosocomial infections one step further by including any clinician-induced illness, infectious or otherwise, regardless of where it was acquired.

EXAMPLE 13.1: WHAT ARE THE POSSIBLE COMPLICATIONS ASSOCIATED WITH DENTAL PROCEDURES IN CATS?

Background: Penetrating ocular trauma during routine dentistry in dogs and cats is an uncommon but devastating iatrogenic post-operative complication that carries a poor prognosis for the eye. The prevalence of periodontal disease in cats has been reported to be up to 98.2%, and hence routine periodontal treatment is essential to managing overall patient health, especially in geriatric cat populations. However, cats may have a higher risk of sustaining ocular trauma during periodontal treatment compared to other species due to their relatively large globe size, the close proximity of the orbit to the oral cavity, and the lack of a bony barrier between the two.

Objectives: Volk et al. (2019) described case signalment along with the dental procedures undertaken in affected cats so that practitioners might mitigate the risks for future cases.

Study design: Case series.

TABLE 13.1
Summary of the Signalment and Outcomes for 13 Cats That Presented with Endophthalmitis Following Tooth Extraction, with or without a Transoral Maxillary Nerve Block, between December 2014 and February 2018

Case	Breed	Sex	Age (Years)	Weight (kg)	Nerve Block	Extractions[a]	Eyes Affected	Onset (Days)[b]	Outcome	Follow-up (Days)[c]
1	DSH	FN	21	3.2	Yes	204, 304	Left	5	Euthanasia	16
2	DSH	FN	13	3.8	Yes	204	Left	1	Enucleation	18
3	Burmese	FN	15	2.7	Yes	204	Left	1	Managed medically[d]	270
4	DSH	FN	16	5.5	Yes	Unknown	Right	0	Enucleation	3
5	Ragdoll	FN	6	4	None	108, 208	Right	0	Enucleation	60
6	DMH	FN	15	4.9	None	107	Right	1	Enucleation	5
7	DMH	FN	6.5	7.5	Not specified	Multiple (8)	Right	3	Enucleation	7
8	DSH	FN	17	3.8	None	Multiple (11)	Left	0	Enucleation	60
9	Himalayan	FN	8.5	3.1	None	208	Left	2	Managed medically[e]	730
10	DLH	FN	10	4	Yes	109, 209	Right	3	Managed medically[e,f]	30
11	DLH	FN	14	3.5	Yes	109	Right	2	Managed medically[e]	60
12	DSH	MN	3	6	Yes	Multiple (8)	Left	1	Enucleation	37
13	DSH	FN	3.5	4	Yes	Multiple (13)	Bilateral	7,14	Managed medically[e]	14 and 365

Source: Volk HA et al. *N Z Vet J* 2019;67:46–51. With permission.
Abbreviations: DLH, Domestic long haired; DMH, Domestic medium haired; DSH, Domestic short haired; FN, female neutered; MN, male neutered.
[a] Tooth/teeth extracted, indicated where known by the modified Triadan system, where the first number represents the oral quadrant (1 = right maxilla, 2 = left maxilla, 3 = left mandible, 4 = right mandible) and the last two numbers represent the tooth (01–03 = incisors, 04 = canine, 06–08 = premolars, 09 = molars).
[b] Interval from dentistry to identification of the signs of endophthalmitis.
[c] Interval from dentistry to the last ophthalmic assessment.
[d] Medical therapy ceased, but evidence of ocular damage persisted.
[e] Active inflammation apparent that was not completely resolved at the last assessment.
[f] Lost to follow-up, subjected to euthanasia 450 days after the dental procedure for non-ocular reasons.

Methods: Medical records from four referral veterinary ophthalmology clinics in Australia and New Zealand were reviewed to identify cats that developed endophthalmitis following dental procedures between December 2014 and February 2018. Thirteen cats were identified (Table 13.1). All cats had at least one maxillary tooth extracted, and some received a transoral maxillary nerve block.

Results: Ocular signs were identified at a median of 1.5 (min 0, max 14) days following elective dental procedures and included fibrin in the anterior chamber, aqueous flare, vision loss, and miosis. Response to medical management was poor overall, with 7/13 (54%) cats undergoing subsequent enucleation, and one cat was euthanatized shortly after the dental procedure due to ocular disease. The remaining five cats were managed medically. Of these, four exhibited signs of persistent inflammation at the last ophthalmic assessment, and one was subsequently euthanatized due to reasons unrelated to ocular health. Active inflammation resolved in one cat; however, the lesions caused by previous inflammation persisted.

Conclusions and Significance: This case series demonstrates that globe penetration during dental procedures carries a poor prognosis for the eye. Clinicians should be aware of the risks

of ocular trauma during dental procedures in cats and great care should be taken to avoid ocular penetration, particularly during tooth extractions. Transoral maxillary nerve blocks should be avoided or used with extreme caution in cats.

FOLLOW-UP QUESTION 13.1

Based on the data summarized in Table 13.1, can the risk of a cat experiencing iatrogenic ocular trauma during routine dentistry be estimated from this study? See Answer 13.1 at the end of this chapter.

13.1.2 Animal Reservoirs

Animal reservoirs of disease agents include (1) carrier animals, e.g., animals (and human beings) with inapparent infections that are also transmitters (or potential transmitters) of an infectious agent, and (2) intermediate hosts and vectors. **Amplifying hosts** may play a role in creating conditions favorable for epidemics of a disease by increasing the abundance of a disease agent in the vicinity of susceptibles.

Animals that have been exposed to an agent may become carriers. **Incubatory carriers** may serve as a source of infection while incubating the disease. This is a characteristic of many viral respiratory infections. **Convalescent carriers** continue to shed infectious organisms after recovering from disease signs and symptoms. This is seen with many parasitic infections caused by protozoa and helminths. Being a carrier does not necessarily mean that an animal is a **reservoir** of infection for others. The pathogen density or location in the carrier may preclude efficient transfer to susceptibles.

Technically, zoonoses are infectious diseases that are spread between animals and people. We tend to think of them as diseases that are transmitted from animals to people, e.g., **anthropozoonoses**. However, in some cases, human beings may be an important source of infection for other animals. These diseases are referred to as **zooanthroponoses**. Examples include methicillin-resistant *Staphylococcus aureus*, influenza A virus, *Cryptosporidium parvum*, *Ascaris lumbricoides*, and cysticercosis of cattle and swine. These terms are not in common use (and can be hard to pronounce), but they do serve to illustrate that disease transmission between people and other animals can go either way (Messenger et al., 2014).

EXAMPLE 13.2: WHAT IS A "CYSTICERCOSIS STORM?"

Background: Human beings are the only definitive hosts of ***Taenia saginata*** and are the source of infection for cattle, the intermediate host. The disease in cattle is known as bovine cysticercosis or, colloquially, "beef measles." Infection in the bovine intermediate host manifests as small, blister-like cystic lesions approximately 5–10 mm in length, occurring most commonly in cardiac, tongue, diaphragm, and/or masseter musculature, but muscles throughout the body may be infected. Each cyst contains a single parasite scolex that, if consumed by a human in raw or inadequately cooked beef, results in infection with a tapeworm, commonly around 4 m in length. The parasite occurs worldwide. Transmission of *T. saginata* relies on humans (definitive hosts) eating cattle (intermediate hosts). The human health effects of infection with the tapeworm stage of *T. saginata* are usually minimal.

In developing countries, cattle usually become infected through accidental ingestion of *T. saginata* eggs while grazing pasture contaminated with human feces through indiscriminate defecation. In more developed countries, sporadic outbreaks are the result

of pasture contamination from raw sewage passing onto pasture as a result of a sewage treatment facility malfunction, particularly on-farm domestic septic tanks or feedlot workers defecating in feed silos, irrigation ditches, and hay paddocks. Sudden exposure of immunologically naïve cattle to the eggs of *T. saginata* in countries where the parasite occurs uncommonly in cattle can lead to a "**cysticercosis storm,**" where an unusually high percentage of cattle are affected

Objectives: Jenkins et al. (2013) sought to determine the source of a reported "cysticercosis storm" among feedlot cattle in north-west New South Wales (NSW), Australia.

Study design: Cross-sectional (prevalence) survey.

Methods: Cystic lesions were detected in the muscles of cattle during routine meat inspection at slaughter. These lesions were confirmed to be cysticerci of *T. saginata* through histological sections of muscle lesions and polymerase chain reaction (PCR) testing using DNA extracted from cysticerci. Data on cattle maintenance were obtained through interviews with feedlot owners and staff. A suspect feed supplement, copra meal, imported from Papua New Guinea was investigated as a possible source of infection. Copra meal is a by-product of the commercial process for extracting oil (coconut oil) from the white coconut flesh.

Results: Between July 5 to December 13, 2010, 390 feedlot cattle from north-western NSW were slaughtered in abattoirs in NSW and Queensland. Of these, 138 animals had been maintained exclusively in feedlot enclosures from 80 to 300 days. Bovine cysticercosis was discovered in 80 cattle (58%; 26 carcasses were condemned). None of 234 animals grazed exclusively on pasture on the property were infected (Table 13.2). No eggs of *T. saginata* were recovered from the copra meal feed supplement using a standard flotation method. Fecal samples from the staff working on the farm were not collected or examined for the presence

TABLE 13.2

Prevalence of *Taenia saginata* Detected in Feedlot Cattle at Slaughter

Slaughter Date	Age (Months)	Weight (kg)	Days on Feed	No. Slaughtered	No. Infected (%)
		Feedlot Cattle			
5 July 2010 (abattoir A)[a]	21–27	560–580	100	26	26 (100)
13 July (abattoir A)	21–27	420–480	80–90	20	14 (70)
13 July (abattoir B)	21–33	>500	180–360	8	5 (62.5)[c]
14 July (abattoir B)[b]	33–>39	>500	180–360	36	21 (58.3)[c]
28 July (abattoir A)	21–27	560–580	100	6	1 (16.6)
20 Aug (abattoir A)	21–27	560–580	100	14	9 (64.2)
28 Sep (abattoir A)	21–27	560–580	100	8	2 (25)
13 Dec (abattoir A)	21–27	420–480	80–90	20	2 (10)
Total				138	80 (58)
		Grass-fed Cattle			
5 July–23 Dec 2010	Various			252	0
		Overall Total		390	80 (21)

Source: Jenkins DJ et al. *Aust Vet J* 2013;91:89–93. With permission.

[a] Abattoir A located in New South Wales.

[b] Abattoir B located in Queensland.

[c] Carcasses were condemned.

of helminth eggs. None of the people associated with the property reported ever eating raw or undercooked beef nor defecating indiscriminately on the property. Sewage treatment of effluent from the only residence on the property was through a septic tank from which the outflow discharged downhill below the feedlot enclosures and not onto cattle pasture. Sewage had never been used to irrigate or fertilize any part of the property, and the property was not subject to flooding.

Conclusions and Significance: This outbreak of bovine cysticercosis was confined to animals fed on feedlot rations. The parasite was not detected in animals exclusively pasture fed in paddocks adjacent to the feedlot enclosures. Therefore, the source of the infection, in the authors' opinion, had to be a component of the feedlot rations. The source of infection most likely arose from rations contaminated with human feces. The suspected source of infection was imported copra meal, which was used as a feed supplement. However, a possible confounding factor for this theory is that there are no reports of *T. saginata* infection in the population of Papua New Guinea, where the copra meal originated.

FOLLOW-UP QUESTION 13.2

The actual source of *T. saginata* in this outbreak was never confirmed, and the circumstantial evidence incriminating copra meal from Papua New Guinea is weak. What additional investigations could have been pursued to rule in or out other sources of infection? See Answer 13.2 at the end of this chapter.

The importance of invertebrate vectors versus vertebrate hosts as reservoirs of disease agents depends on the lifespan of the respective hosts and the survival and infectivity of the disease agent in their tissues. Experimental studies may provide important information directly applicable to field situations.

Amplifying hosts are generally considered to be those intermediate hosts that do not suffer from disease, but in which the number of infectious units increases extensively, thus providing a source of infection for epidemics in humans or domestic animals. **West Nile virus (WNV)** provides an excellent example of the role of amplifying hosts.

EXAMPLE 13.3: HOW CAN THE RELATIVE IMPORTANCE OF AVIAN AMPLIFYING HOSTS FOR WEST NILE VIRUS OUTBREAKS BE DETERMINED?

Background: WNV is a mosquito-borne arbovirus known for outbreaks of neurologic disease and death among people, horses, and birds in temperate regions of North America. It is the leading cause of mosquito-borne human disease in the continental United States (CDC, 2018). The basic cycle of the infection involves wild birds and ornithophilic mosquitoes. Peridomestic birds serve as amplifiers of the virus, but their reservoir competence varies widely.

Objectives: Komar et al. (2013) sought to identify and assess the relative importance of avian amplifying hosts of WNV in the Phoenix, Arizona, area where WNV has caused sporadic outbreak of human encephalitis since 2003.

Study design: Cross-sectional (prevalence) survey.

Methods: Blood samples were drawn from 300 (representing 17 species) of resident birds over a 2-week period in the fall following an outbreak in the East Valley of metropolitan

TABLE 13.3

Calculation of the Relative Number of West Nile Virus (WNV) Infections and the Modified Mosquito Inoculation Index among Four Abundant Species of Birds in the Study Area in the East Valley of Maricopa County, AZ

Species	Relative Abundance[a] A (95% CI)	WNV Antibody Prevalence S (95% CI)	Estimated Relative No. of Infections = A × S(95% CI)	Competence C (95% CI)	Mosquito Inoculation Index = A × S² × C(95% CI)
Mourning dove[b]	63.9 (60.5–67.3)	0.35 (0.24–0.48)	22.5 (16.2–33.4)	0.096 (0.012–0.18)	0.8 (0.3–7.1)
House finch	8.8 (3.8–13.8)	1.00 (0.57–1.00)	8.8 (3.8–18.2)	1.19 (0.77–1.61)	10.4 (4.3–24.4)
House sparrow	11.4 (9.1–13.6)	0.51 (0.42–0.61)	5.9 (4.4–7.7)	1.12 (0.43–1.81)	3.4 (1.8–9.7)
Great-tailed grackle	13.3 (8.8–17.8)	0.86 (0.49–0.97)	11.4 (7.4–21.5)	1.28 (0.49–2.07)	12.5 (2.8–24.8)

Source: Komar N et al. *Am J Trop Med Hyg* 2013;89:474–481. With permission.

Abbreviation: CI = confidence interval.

[a] Calculated as birds per party-hr.

[b] Only infection rate for hatch-year mourning doves is used for this analysis because statistical analysis infers that numerous adult doves were already seropositive from transmission in previous years. For the other species, the infection rate for all ages is used.

Phoenix during summer, 2010, and WNV antibody prevalence determined. WNV antibody was detected in 144 samples (48%) from 14 species.

Results: House sparrows (*Passer domesticus*), house finches (*Haemorhous mexicanus*), great-tailed grackles (*Quiscalus mexicanus*), and mourning doves (*Zenaida macroura*) accounted for most WNV infections among locally resident birds (Table 13.3). These species roost communally after early summer breeding. In September 2010, *Culex* vector-avian host contact was threefold greater at communal bird roosts compared with control sites, as determined by densities of resting mosquitoes with previous vertebrate contact (i.e., blood-engorged or gravid mosquitoes).

Conclusions and Significance: Because of the low reservoir competence of mourning doves, this species was considered a weak amplifier but a potentially useful free-ranging sentinel. Highly competent sparrows, finches, and grackles were predicted to be key amplifying hosts for WNV in suburban Phoenix. That, together with increased density of the human population, creates the conditions necessary for future epidemics.

FOLLOW-UP QUESTION 13.3

What aspects of this study might influence the interpretation of results? See Answer 13.3 at the end of this chapter.

13.1.3 Environment

The environment may be considered a source of infection when the disease agent multiplies there, not requiring any animal host for its continued survival. *Histoplasma capsulatum*, causative agent of histoplasmosis, is an example of an infectious, non-transmissible disease agent with an

environmental source. Infection results from inhalation of airborne conidia that are produced during growth of organisms in the soil. See the next section for a further discussion of transmissible versus non-transmissible diseases.

During the course of an outbreak investigation, a distinction should be made between those situations in which the environment is the ultimate source and reservoir of infection and those in which the environment is a **fomite** or vehicle of transmission. In the latter case, even though the immediate source of a disease agent, such as parasite ova in the soil, is environmental, the ultimate source of infection is another host.

13.2 TRANSMISSION

13.2.1 Mode of Transmission versus Route of Infection

A distinction must be made between the terms **mode of transmission** and **route of infection**. For example, if we say that the mode of transmission is via the respiratory tract, we have not indicated whether the organisms gained access via droplet transmission (direct), droplet nuclei, or dust (airborne). The respiratory tract is really the route of infection. The **mode of transmission** refers to the way(s) in which an etiologic agent is transmitted from affected to susceptible individuals.

Modes of transmission may be broadly classified as horizontal or vertical, and within horizontal as direct, indirect, or airborne. Routes of infection (and exit) include alimentary, respiratory, urogenital, anal, skin, and conjunctival.

Modes of transmission may be broadly classified as horizontal or vertical, and within horizontal as direct, indirect, or airborne. The **route of infection** refers to the route by which an etiologic agent gains access to the body of a susceptible individual. Routes of infection (and exit) include alimentary, respiratory, urogenital, anal, skin, and conjunctival (Figure 13.1). **Mode of spread** or dissemination refers to how a disease agent is spread from one individual or geographic region to another.

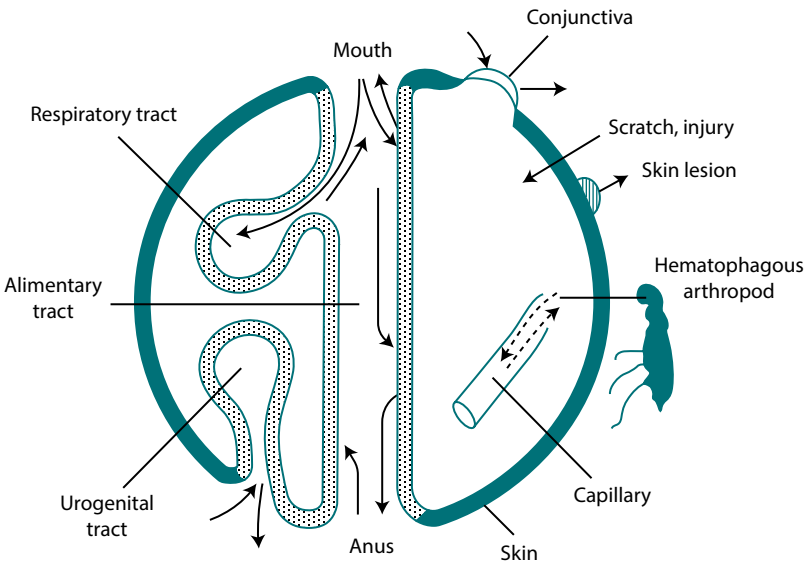

FIGURE 13.1 Diagram illustrating the routes of exit and entry of infectious agents in vertebrate animals. (With kind permission from Springer Science+Business Media: *Population Biology of Infectious Diseases*, 1982, Anderson RM and May RM.)

13.2.2 Transmissible versus Non-Transmissible Diseases

Diseases are broadly classified as **transmissible** (communicable) or **non-transmissible**. Transmissible disease may be due to a specific infectious agent or its toxic products (such as the carcass-maggot cycle of waterfowl botulism), which may arise through transmission of that agent or its products from a reservoir to a susceptible host. Transmission may occur directly, as from an infected person or animal, or indirectly through an intermediate plant or animal host, vector, or the inanimate environment.

Non-transmissible diseases may be caused by infectious or non-infectious agents. Infectious agents may originate from environmental sources (such as the saprophytic fungi responsible for histoplasmosis, blastomycosis, and coccidioidomycosis, or infections caused by *Clostridium tetani*), or part of the normal flora such as the bacterial secondary invaders responsible for pneumonia, wound infections, and abscesses. Non-infectious agents include poisons and environmental toxins, immunologic and metabolic mechanisms, nutritional deficiencies, and functional defects (such as congenital anomalies).

The term **contagious** is used to describe diseases that are transmissible from an infected to an uninfected individual, typically by direct contact. The term often appears in association with many diseases of veterinary interest such as contagious pleuropneumonia, contagious ecthyma, contagious equine metritis, contagious agalactia, contagious mastitis, contagious digital dermatitis, and others.

> ***Practically speaking, introduction into the herd of an animal afflicted with a non-transmissible disease does not increase the likelihood of disease in others.***

Contact with diseased animals is always viewed with some degree of apprehension. Practically speaking, introduction into the herd of an animal afflicted with a non-transmissible disease does not increase the likelihood of disease in others. Introduction into the herd of an individual with a transmissible disease increases the likelihood of disease for others. The degree of risk depends, in part, on the mode of transmission.

13.3 MODES OF TRANSMISSION

Transmission of an infectious agent may occur **horizontally** between contemporaries or animals of more or less the same generation directly, indirectly, or via airborne routes. Transmission may also occur **vertically** by transmission from infected animals of one generation to animals of the succeeding generation (*in utero* or via colostrum). The modes of transmission of disease agents are depicted in Figure 13.2 and described below.

13.3.1 Horizontal Transmission

Horizontal transmission describes the transmission of a disease agent among contemporaries. Modes of horizontal transmission may be direct, indirect, or airborne.

13.3.1.1 Direct Transmission

Direct transmission implies direct and essentially immediate transfer of an agent from infected to susceptible hosts. This may occur by **direct contact**, as through touch, a scratch, lick, bite, or intercourse, or through **direct projection**, where atomized droplets are sprayed onto the conjunctiva or mucous membranes of the eye, nose, or mouth during coughing or sneezing. Direct projection, also known as **droplet spread**, is usually limited to a distance of 1 meter or less.

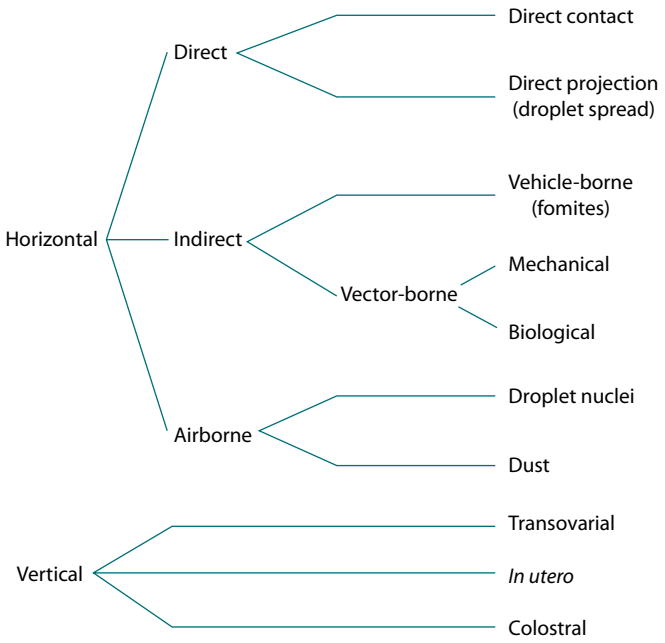

FIGURE 13.2 Modes of transmission of disease agents.

Direct transmission implies direct and essentially immediate transfer of an agent from infected to susceptible hosts. Indirect transmission implies the passage of infectious agents between individuals through the medium of inanimate or animate objects.

13.3.1.2 Indirect Transmission

Indirect transmission implies the passage of infectious agents between individuals through the medium of inanimate or animate objects. The time period between contamination of the object and subsequent exposure of susceptible individuals is highly variable and may range from a few minutes to years. Indirect transmission may be vehicle borne or vector borne. Most parasitic diseases are transmitted indirectly, either from environmental contamination or via intermediate hosts.

13.3.1.2.1 Vehicle-Borne Transmission

Vehicle-borne transmission occurs through exposure to contaminated inanimate objects (**fomites**) such as bedding, surgical instruments, soil, water, food, milk, and biological products (including blood, serum, plasma, tissues, or organs). The agent may or may not have multiplied or developed in or on the vehicle before being transmitted. The term **fomite** originates from the Latin word for tinder, *fomes*. The equipment of sick animals has long been thought of as forms of smoldering tinder, which can "ignite" the fire of disease in others.

The equipment of sick animals has long been thought of as forms of smoldering tinder, which can "ignite" the fire of disease in others.

13.3.1.2.2 Vector-Borne Transmission

Vector-borne transmission is generally understood to mean transmission by invertebrate vectors, such as flies, mosquitoes, or ticks. In some cases, vertebrate hosts such as dogs, foxes, or bats may serve as vectors, as in the case of rabies transmission. Transmission may be by injection of salivary gland fluid during biting or by regurgitation or deposition on the skin of feces or other body fluids that contaminate host tissues through the bite wound or through an area of trauma induced by scratching or rubbing. Vector-borne transmission may be either mechanical or biological.

Mechanical transmission results from simple mechanical carriage of the disease agent between hosts by crawling or flying arthropods. It does not require multiplication or development of the disease agent in the vector. The disease agent is transmitted between hosts on soiled appendages or the proboscis, or by passage of organisms through the gastrointestinal tract.

Biological transmission requires a period of multiplication, cyclic development or both before the vector can transmit the infective form of the agent. This period is referred to as the **extrinsic incubation period** as opposed to the **intrinsic incubation period** required for an infected vertebrate host to become infective. The disease agent may be transmitted vertically (**transovarially**) between generations of the vector or **transstadially** from one stage to another within a single generation.

> ***Horizontal transmission describes the transmission of a disease agent among contemporaries. Vertical transmission describes the transmission of a disease agent from animals of one generation to subsequent generations.***

13.3.1.3 Airborne Transmission

Airborne transmission involves the dissemination of microbial aerosols. **Microbial aerosols** are suspensions of particles in the air consisting partially or wholly of microorganisms. They may remain suspended in the air for long periods of time and usually infect the host via the respiratory tract. Particle diameters range from less than 1 to 100 μm. Droplets and other large particles that promptly settle out of the air are not considered airborne. Airborne transmission may be accomplished by droplet nuclei or dust.

Droplet nuclei are the small residues that result from evaporation of fluid from droplets emitted by an infected host. They may also be created by atomizing devices, accidentally in microbiology laboratories, abattoirs, rendering plants, or necropsy rooms. Droplet nuclei usually remain suspended in the air for long periods of time. **Dust** consists of the small particles of widely varying size that may arise from soil (as fungus spores separated from dry soil by wind or mechanical agitation), clothes, bedding, or contaminated floors.

13.3.2 Vertical Transmission

Vertical transmission describes the transmission of a disease agent from animals of one generation to subsequent generations. Vertical transmission may be transovarial, e.g., between generations of invertebrate vectors via the egg, *in utero*, or transplacental, e.g., from parent to offspring within the uterus, or colostral from parent to offspring at parturition via colostrum or milk. Vertical transmission provides an important reservoir or overwintering mechanism for certain vector-borne viruses, rickettsia, and protozoa.

EXAMPLE 13.4: WHAT OCCUPATIONAL HAZARDS DO VETERINARIANS AND VETERINARY PERSONNEL FACE IN PRACTICE?

Background: Zoonotic diseases are recognized occupational hazards faced by veterinary personnel on a daily basis (Williams et al., 2015). Of the approximately 1415 known human pathogens 868 (61%) are reported to be zoonotic, and 132 of 175 (75%) emerging diseases that affect humans are zoonotic. Veterinary, medical, and public health personnel recently revised and updated the Compendium of Veterinary Standard Precautions (VSP) for Zoonotic Disease Prevention in Veterinary Personnel (https://avmajournals.avma.org/doi/10.2460/javma.247.11.1252).

Objectives: The objectives of the compendium are to address infection prevention and control issues specific to veterinary practice; provide practical, science-based veterinary infection control guidance; and provide a model infection control plan for use in individual veterinary facilities.

Study design: Literature review.

Methods: This 2015 version of the compendium was extensively revised and updated since the 2010 version based on reviews of the human and veterinary medical literature and the input of a panel of consultants.

Results: A perusal of Appendix 1 of the full report is instructive, as it summarizes the myriad ways in which more than 50 zoonotic diseases of importance in the United States can be acquired and transmitted. During their careers, approximately two-thirds of veterinarians report a major animal-related injury that resulted in lost work time or hospitalization. The most common occupational injuries among veterinary personnel include animal bites and scratches, kick and crush injuries, and needlesticks.

Conclusions and Significance: Although it may not be possible to eliminate all zoonotic disease hazards, employers should conduct a workplace risk assessment and implement appropriate control measures where possible. Adherence to a well-developed employee safety and health program will minimize the risk of injury and illness. The compendium provides reasonable guidance for minimizing one type of workplace hazard—zoonotic disease transmission—among veterinary personnel in clinical settings through the application of the VSP.

FOLLOW-UP QUESTION 13.4

Given the variety of zoonotic agents. hosts, and modes of transmission listed in the VSP, what kinds of veterinary procedures increase the risk of exposure of veterinarians and their staff to zoonotic diseases? See Answer 13.4 at the end of this chapter.

13.4 FACTORS AFFECTING COMMUNICABILITY

Communicability may be defined as the ease with which a disease agent is spread within a population. One way of expressing communicability is the **basic** (or **intrinsic**) **reproduction number** ($\mathbf{R_0}$), which represents the average number of secondary infections generated by one primary case in a susceptible population, and can be used to estimate the level of immunization or other risk reduction strategy required to control an epidemic (Delamater et al., 2019). It follows that as herd immunity (see Chapter 12) increases, the number of secondary cases declines by a factor roughly proportional

to the fraction that is susceptible. This relationship is described by the equation $R = (1 - P) \times R_o$ where R = the **effective reproductive number** and P = the proportion of the population immune or resistant, i.e., herd immunity. If $R > 1$, then the number of infected individuals can increase, possibly leading to an epidemic. If R approximates 1, then conditions supporting endemic disease exist. If $R < 1$, then disease frequency will decline, possibly leading to eradication.

Communicability may be defined as the ease with which a disease agent is spread within a population. One way of expressing communicability is the basic (or intrinsic) reproductive number (R_o), which represents the average number of secondary infections generated by one primary case in a susceptible population, and can be used to estimate the level of immunization or other risk reduction strategy required to control an epidemic.

The value of R_o depends on three parameters: (1) the duration of the infectious period, (2) the probability that a contact between an infective and a susceptible individual will lead to an infection, and (3) the number of new susceptible individuals contacted per unit time (Delamater et al., 2019). The communicability of a disease agent is also determined by factors that are specific to the disease agent, its hosts, and the environment, e.g., the **agent-host-environment triad**. Some of these factors are discussed in the following sections.

13.4.1 Agent Factors

13.4.1.1 Life Cycle

The life cycle of a disease agent may be defined as the sequence of developmental stages from infection of one host to infection of a second host. Epidemiologically, the life cycle can be expressed as discrete time periods. Included are the prepatent period, communicable period, and extrinsic incubation period.

The **prepatent period**, or **intrinsic incubation period**, is the time between infection of the vertebrate host and detectability of an agent in secretions, excretions, blood, or tissues. The **communicable period** is the time or times during which an infectious agent may be transferred directly or indirectly from one infected animal to another, including invertebrate vectors. The **extrinsic incubation period** is the period of time between infection of a biological vector and acquisition by the vector of the ability to transmit the agent to another susceptible vertebrate host. The extrinsic incubation period is a major determinant of the time between introduction of an infectious animal into a herd and occurrence of disease among susceptibles.

13.4.1.2 Minimal Infective Dose

Disease agents vary widely in their infectivity for a host. Generally speaking, the lower the **minimal infective dose**, the more readily the agent is transmitted.

13.4.2 Host Factors

13.4.2.1 Heterogeneity

Within any population, individuals vary in their susceptibility to infection and disease, irrespective of their immune status. This phenomenon, generally referred to as **innate resistance**, is most likely an expression of the genetic composition of the host. By limiting infection, transmission is reduced. On the contrary, certain individuals may be particularly susceptible to infection and serve as a reservoir of infection for the rest of the herd. For example, the term *lousy* refers to the propensity of certain individuals to develop heavy louse infestations, particularly in the winter and spring (Benelli et al., 2018).

13.4.2.2 Immunity

Generally, vertebrate hosts develop a stronger immune response to microbial pathogens than they do to metazoans. This may be a result of the extensive multiplication of the former in the host and the associated strong antigenic exposure. As a result, microbial infections tend to be of shorter duration and self limiting, thus limiting the opportunity for secondary transmission.

13.4.3 Environmental Factors

13.4.3.1 Particle Diameter

13.4.3.1.1 Droplets

The efficiency of transmission by direct projection is limited by the size of the **droplets**, which are greater than 100 μm in diameter. The typical settling velocity of the droplets is greater than 1 foot per second, and the time of suspension is less than 3 seconds. Their flight range is restricted to 1 to 2 meters or less (Williams et al., 2015). Droplet spread can be effectively reduced through use of a face mask and by reducing crowding among animals.

13.4.3.1.2 Dust Particles

Dust particles are smaller than droplets, ranging from 10 to 100 μm in diameter. Their suspension time is limited by their settling velocity, which ranges from 1 foot per minute to 1 foot per second. They typically hover in clouds and can be removed from the air by filtration and electrostatic precipitation. Dust-borne spread can be reduced by air cleanliness and moistening or oiling contaminated sources.

13.4.3.1.3 Droplet Nuclei

Droplet nuclei are the smallest of the particles, ranging from 2 to 10 μm in diameter. Their settling velocity is less than 1 foot per minute. They are most efficiently dispersed throughout confined atmospheres, as in hog houses or abattoirs, and their time of suspension is limited indoors by the degree of ventilation. They can be removed from the air by electrostatic precipitation, and droplet spread can be reduced through sanitary ventilation, e.g., air change and equivalent air disinfection.

13.4.3.2 Microclimate

There is strong evidence to suggest that climate change has and will continue to affect the occurrence, distribution, and prevalence of livestock diseases globally. The effects of climate change are mediated through a variety of interdependent factors, including the molecular biology of the pathogen, vectors, farming practice and land use, and zoological and environmental factors together with the establishment of new microhabitats. Gale et al. (2009) have proposed a risk assessment framework that accommodates these factors and could be used to screen for the emergence of unexpected disease events in Great Britain. Central to the risk assessment is the identification of those factors through which climate change could affect livestock diseases. This approach should be generalizable to animal diseases worldwide, including diseases of wildlife.

REFERENCES

Anderson RM and May RM. *Population Biology of Infectious Diseases*. New York: Springer-Verlag; 1982.

Benelli G, Caselli A, Di Giuseppe G et al. Control of biting lice, Mallophaga—A review. *Acta Trop* 2018;177:211–219.

CDC, 2018. West Nile Virus. U.S. Centers for Disease Control and Prevention. Available online at: https://www.cdc.gov/westnile/index.html

Delamater PL, Street EJ, Leslie TF et al. Complexity of the basic reproduction number (R_0). *Emerg Infect Dis* 2019;25:1–4. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6302597/

Gale P, Drew T, Phipps LP et al. The effect of climate change on the occurrence and prevalence of livestock diseases in Great Britain: A review. *J Appl Microbiol* 2009;106:1409–1423. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/j.1365-2672.2008.04036.x

Jenkins DJ, Brown GK, Traub RJ. "Cysticercosis storm" in feedlot cattle in north-west New South Wales. *Aust Vet J* 2013;91:89–93.

Komar N, Panella NA, Young GR et al. Avian hosts of West Nile virus in Arizona. *Am J Trop Med Hyg* 2013;89:474–481. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3771284/

Messenger AM, Barnes AN, Gray GC. Reverse zoonotic disease transmission (zooanthroponosis): A systematic review of seldom-documented human biological threats to animals. *PLOS ONE* 2014;9:e89055. Full text available at: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0089055

Volk HA, Bayley KD, Fiani, N et al. Ophthalmic complications following ocular penetration during routine dentistry in 13 cats. *N Z Vet J* 2019;67:46–51.

Williams CJ, Scheftel JM, Elchos BL et al. Compendium of veterinary standard precautions for zoonotic disease prevention in veterinary personnel: National Association of State Public Health Veterinarians: Veterinary infection control committee. *J Am Vet Med Assoc* 2015;247:1252–1277. Full text available at: https://avmajournals.avma.org/doi/10.2460/javma.247.11.1252.

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 13.1: This was a case series study design and the population-at-risk is unknown. Further, cats presented with a variety of potential predisposing characteristics (age, breed, sex, etc.). Thus, the exploration of risk factors for post-surgical endophthalmitis cannot be determined. The fact that only 13 cats were referred for endophthalmitis over the 39-month study period suggests that this may be a rare, albeit serious, complication of dental procedures in the cat.

Answer 13.2: As employees working on the farm were not tested for *T. saginata* infestation, local contamination of feed sources could not be ruled out.

Answer 13.3: The authors point out that the seroprevalence values observed reflect a wide scale in time and space. In this case, the infections probably took place over a 4–5-month period. Therefore, their exposure status may reflect time spent beyond the outbreak area. The effect of sampling birds that had traveled from beyond the outbreak area will dilute the apparent importance of that species as an amplifying host. A second limitation of the retrospective mosquito inoculation index calculation was the use of seroprevalence as a surrogate for the avian infection rate. If WNV infections also resulted in undetected avian mortality, then the infection rates would be underestimated, and the estimated importance of the affected species as an amplifier also underestimated.

Answer 13.4: The following exposure risks were highlighted in the VSP report:

- Hand-to-mouth direct or indirect transmission during examination, treatment or handling of animals or contaminated objects.
- Droplet transmission created by coughing, sneezing, or vocalization of patients; lancing abscesses; or dentistry.
- Airborne transmission during suction of bronchoscopy, the use of power tools during necropsy, and during high-pressure spraying of equipment and work in confined animal housing units.
- Vector-borne transmission through the introduction of flea or tick vectors into the clinic, or from exposure during field work.
- Necropsy is a high-risk procedure because of the possibility of injury and potential contact with infectious agents in body tissues, body fluids, and aerosols.

14 The Cost of Disease

14.1 DEFINING DISEASE IN ECONOMIC TERMS

Earlier in the text, we discussed how disease could be defined in a variety of ways, including animal performance. A producer's decision as to whether to institute any sort of disease control program will be based, in large part, on economic considerations. Similarly, the relative merits of alternative regional or national disease control strategies are usually evaluated on the basis of expected short- and long-term economic impacts.

In order to better target a disease control program, some sort of economic analysis is usually necessary. A variety of economic modeling approaches have been used in veterinary medicine. Partial budgeting, benefit-cost analysis, and decision analysis are among the most common (Rushton, 2009). In the next section, the "measures of effect" approach is used to introduce the topic and illustrate how the relative importance of risk factors can be compared in economic terms. Subsequent sections use more complex models to evaluate disease control programs based on their benefits and costs.

14.1.1 The "Measures of Effect" Approach to Estimating Disease Impact

The following example takes advantage of the concept of "**measures of effect"** for expressing risk that was introduced in Chapter 6. In this case, risk is expressed in economic terms to determine which risk factors have the greatest economic impact.

A producer's decision as to whether to institute any sort of disease control program will be based, in large part, on economic considerations.

EXAMPLE 14.1: DO MEDICAL COMPLICATIONS AT CALVING INFLUENCE SUBSEQUENT REPRODUCTIVE PERFORMANCE OF DAIRY COWS?

Background: Reproductive performance is crucially important to maintain profitability in the dairy industry. In Korea, continuous breeding for an increased milk yield and use of intensive production systems have dramatically increased milk production per cow but decreased reproductive performance. Conception at the first service after calving is key to optimal reproductive performance in dairy cows, although the percentage success of first service has been shown to range between 26.7% and 50.7%. A failure of first-service conception (FSC) may lead to an increase in the number of days open, insemination numbers, reproductive treatment, feeding, culling, and replacement heifers. Thus, the identification of factors that potentially limit the success of FSC, including biological and environmental conditions, might be useful to improve reproductive performance in herds with high yields under intensive production systems.

Objectives: Kim and Jeong (2019) sought to identify and assess the economic impact of potential risk factors limiting the first-service conception rate in dairy cows in Korea.

Study Design: Cohort study.

Methods: Data were collected from 790 lactations (340 primiparous and 450 multiparous) in 426 cows regarding cow parity, peri- and postpartum disorders, body condition score (BCS), reproductive performance, and expenses associated with reproductive management (treatment, culling, and others) on two dairy farms from 2011 to 2016. Initially, risk factors limiting FSC rate in dairy cows were evaluated by univariate and multiple logistic regression. Odds ratios and 95% confidence intervals were determined by logistic regression. Various biological and environmental factors, such as herd, cow parity, BCS at 1 month postpartum and first artificial insemination (AI), resumption of cyclicity within 1 month of calving, year, AI season, insemination at detected estrus or timed AI, pre-existence of peri- and postpartum disorders, and calving to first AI interval, were considered (Table 14.1). A p-value <0.05 was

TABLE 14.1
Risk Factors Evaluated for Their Influence on First-Service Conception Rate in Dairy Cows

		No. of Cows		
Variable	**Level**	**AI**	**Conceived**	**%**
Farm	A	350	157	44.9
	B	440	177	40.2
Cow parity	1	243	109	44.9
	2	206	85	41.3
	3	153	67	43.8
	≥4	188	73	38.8
BCS at 1 month postpartum	<2.75	126	47	37.3
	≥2.75	664	287	43.2
CL detection within 1 month postpartum	No	467	202	43.3
	Yes	323	132	40.9
Year	2011–2012	208	98	47.1
	2013–2014	294	119	40.5
	2015–2016	288	117	40.6
AI season	Spring	177	88	49.7
	Summer	156	45	28.8
	Autumn	255	105	41.2
	Winter	202	96	47.5
Timed AI	No[a]	594	258	43.4
	Yes	196	76	38.8
Peri- and postpartum disorders[b]	No	526	249	47.3
	Yes	264	85	32.2
Calving to first AI interval (days)	<80	347	143	41.2
	≥80	443	191	43.1
BCS at first AI	<3.0	183	62	33.9
	≥3.0	607	272	44.8

Source: Kim I-H, and Jeong JK. *Asian-Australas J Anim Sci* 2019;32:519–526. With permission.
Abbreviations: AI, artificial insemination; BCS, body condition score; CI, confidence interval.
[a] Insemination at detected estrus (IDE).
[b] Peri- and postpartum disorders requiring veterinary intervention include dystocia, retained placenta, septicemic metritis, clinical endometritis, ketosis, milk fever, and abomasal displacement.

considered statistically significant. Next, the economic impact of the success or failure of FSC was evaluated by comparing the expense associated with reproductive management until conception between cows that did or did not conceive at their first service.

Results: Cows with BCS <3.0 had a lower probability of conceiving at first insemination (OR = 0.64, $p < 0.05$) than cows with BCS ≥3.0. Cows inseminated during summer were less likely to conceive (OR = 0.44, $p < 0.001$) than cows inseminated during spring. Cows with peri- or postpartum disorders (dystocia, retained placenta, septicemic metritis, clinical endometritis, ketosis, milk fever, and abomasal displacement) requiring veterinary intervention were less likely to conceive (OR = 0.55, $p < 0.001$) than cows without disorders. However, farm, cow parity, BCS at 1 month postpartum, detection of a CL within 1 month of calving, year, timed AI, and calving to first AI interval were not associated with the FSC rate ($p > 0.05$). Cows failing to conceive required additional expenditure on reproductive treatment ($55.40; see Table 14.2) and other management ($567.00) compared to cows that conceived at first insemination. Thus, a total of $622.40 extra was spent on reproductive treatment and other management for cows that failed to conceive at their first AI.

Conclusions and Significance: The authors concluded that lower BCS, hot weather at first insemination, and peri- and postpartum disorders are risk factors limiting FSC, which result in an economic loss of $622.40 per dairy cow. It must be pointed out that the reported economic loss of $622.40 per dairy cow reflects the cost per affected cow, not per cow exposed to the identified risk factors nor as a member of the herd.

Veterinarians are most likely to have been involved in the treatment of peri- and postpartum disorders. The "measures of effect" approach (see Chapter 6) can be used to assess the actual economic impact of peri- and postpartum disorders upon the cost of reproductive management at the herd level. The economic costs of reproductive treatment required to achieve conception in cows that did or did not conceive on first AI service are summarized in Tables 14.2 and 14.3. Economic costs are substituted for disease frequency in the analysis. Table 14.4 summarizes the simple and compared risks, in economic terms, of peri- and postpartum disorders and their impact at the herd level. The population attributable fraction is 8.8%, suggesting that peri- and postpartum disorders do not have a significant influence upon the costs of failure of

TABLE 14.2

Costs of Reproductive Treatment Required to Achieve Conception in Cows that Did or Did Not Conceive at Their First AI

Item	Unit	Value ($)/Dose	Cows that Did Not Conceive at First AI (n = 384)	Cows that Did Conceive at First AI (n = 334)
PGF_{2a}	1 dose	3.5	1.98 doses × $3.5 = $6.93	1.39 doses × $3.5 = $4.87
GnRH	1 dose	2.5	2.14 doses × $2.5 = $5.35	1.49 doses × $2.5 = $3.73
CIDR	1 dose	22	0.49 doses × $22 = $10.78	0.14 doses × $22 = $3.08
Semen	1 straw	20	2.75 straws × $20 = $55.00	1 straw × $20 = $20.00
Palpation	1 time	7	4.36 palpations × $7 = $30.52	3.07 palpations × $7 = $21.50
Total			$108.58	$53.18

Source: Kim I-H, and Jeong JK. *Asian-Australas J Anim Sci* 2019;32:519–526. With permission.

Abbreviations: AI, artificial insemination; PGF2a, prostaglandin F2a; GnRH, gonadotrophin releasing hormone; CIDR, a controlled, internal drug-release device containing 1.9 g progesterone.

TABLE 14.3
Effect of Peri- and Postpartum Disorders[a] upon First-Service Conception Rate and Cost of Reproductive Treatment

Postpartum Disorder (At Risk)	No. of Cows	At Risk (%)	(a) Failed to Conceive (%)	(b) Treatment Cost per Affected Cow	(a × b) Treatment Cost per Cow at Risk
Yes	264	33	67.8	$55.40	$37.56
No	526	67	52.7	$55.40	$29.17
Total	790			Mean =	$31.98[b]

Source: Kim I-H, and Jeong JK. *Asian-Australas J Anim Sci* 2019;32:519–526. With permission.

[a] Peri- and portpartum disorders requiring veterinary intervention include dystocia, retained placenta, septicemic metritis, clinical endometritis, ketosis, milk fever, and abomasal displacement.

[b] Weighted mean cost per at-risk cow based on an additional treatment cost of $55.40 for cows that failed to conceive at their first AI. Data from Table 14.2.

TABLE 14.4
Impact of Peri- and Postpartum Disorders upon Cost of Reproductive Treatment

Simple Risks

Treatment costs in exposed = $37.56
Treatment costs in unexposed = $29.17
Treatment costs/cow overall = $31.98
Prevalence of exposure = 33%

Compared Risks

Relative risk ($37.56 ÷ $29.17) = 1.29
Attributable risk/cow ($37.56 − $29.17) = $8.39
Population attributable risk/cow (0.33 × $8.39) = $2.80
Population attributable fraction ($2.80 ÷ $31.98) = 8.8%

Source: Data from Table 14.3.

first-service conception in the herd. This is probably due to the fact that the difference in first-service conception rates for at-risk versus not-at-risk cows is relatively small (see Table 14.1). Many cows that did not experience peri- or postpartum disorders still failed to conceive at first service. Thus, although the effect of prior calving difficulties upon first-service conception is statistically significant, it may not be economically significant at the enterprise level. Dairy managers and veterinary practitioners can look for other approaches to improving first-service conception rates and reducing the cost of dairy production.

FOLLOW-UP QUESTION 14.1

The analyses in Tables 14.3 and 14.4 incorporate the $55.40 additional expenditure on reproductive treatment of non-conceiving cows, but not the additional $567 for other management costs in their analysis. How would the addition of the additional $567 affect the results? See Answer 14.1 at the end of this chapter.

14.1.2 Partial Budgeting and Benefit-Cost Analysis

14.1.2.1 Partial Budgeting

In order to estimate benefits and costs to producers of a specific disease control program, **partial budget analysis** can be used (Rushton, 2009). The part of the enterprise budget affected by the disease is separated out so that the effect of the disease is not overshadowed by some other factor or disease. Fixed costs (such as labor costs, machinery and building operating costs and depreciation, rent, and interest) are excluded from the analysis.

Partial budgeting usually places farm budget items into one of four categories:

1. ***Additional returns*** due to adoption of a proposed control program.
2. ***Forgone returns*** such as income lost from a reduced number of culled animals.
3. ***Additional costs incurred*** due to the control procedure such as drugs and management procedures.
4. ***Costs no longer incurred*** such as veterinary expenses.

The disease control program should be adopted if the sum of (1) and (4) is greater than that of (2) and (3).

14.1.2.2 Benefit-Cost Analysis

Benefit-cost analysis is a method for estimating the profitability of disease control programs over an extended period of time (Rushton, 2009). There are three main elements involved: (1) enumeration of benefits and costs (as described above for partial budget analysis), (2) selection and application of a discount rate to benefits and costs, and (3) specification of a decision criterion. **Cost-effectiveness analysis** is a variant of benefit-cost analysis that is used when the expected benefits are difficult to quantify in economic terms. A training workshop might be evaluated based on how many participants adopt the new technology. Preference is given to the program that, given its costs, benefits the largest number within the target population.

14.1.2.3 Discounting, Present and Future Value of Money

Veterinarians are familiar with interest rates on investments or loans as an indicator of the **time value of money**. In contrast, the **discount rate** and the process of discounting used in calculating **present values** for a benefit-cost analysis is less familiar. Because benefits and costs of a long-term disease control program do not occur simultaneously, they cannot be compared without adjusting for the time value of money. Further, costs may accrue during the relatively short life of the program. Benefits may accrue indefinitely into the future.

> ***Because benefits and costs of a disease control program do not occur simultaneously, they cannot be compared without adjusting for the time value of money.***

The interest rate determines the value of the principal of an investment at a future date. The discount rate is the reverse of interest rate. If, for example, we were to invest $500 in a disease control program that would yield a $1000 return 5 years from now, the current benefit of the program would not be $1000. This is because $1000 invested today will be worth considerably more than $1000 five years from now. If, for example, we assume a 3% interest rate over the next 5 years, $1000 five years from now would be equivalent to $862.61 invested today.

Using a discount rate, disease control program benefits and costs that accrue in the future are discounted to present values. The formula for calculating future value from present value is:

$$FV = PV\ (1 + r)^n$$

and the formula for calculating present value from future value is:

$$PV = \frac{1}{(1+r)} \times FV$$

where PV = present value, FV = future value (i.e., the value of a benefit or cost), r = discount rate (usually the prevailing interest rate paid by loan institutions), and n = the interest compounding interval, usually expressed in years. As the time (n) before a benefit is received increases, the present value of future benefits decreases.

14.1.2.4 Decision Criteria in Benefit-Cost Analysis

Three measures are commonly used to interpret the results of a benefit-cost analysis and arrive at a decision (Rushton, 2009):

a. *Net Present Value*: The **net present value** expresses the difference between the total present value of benefits and their costs. Stated another way, net present value is the present value of net benefits. It represents the value of the program at today's prices. It indicates the scale of the net benefits but does not show the relative size of the benefits and costs. Expensive programs will tend to have a high NPV, even if the return on investment is small.
b. *Benefit-Cost Ratio (B/C ratio)*: The ratio of total present benefits to costs is the **benefit-cost ratio** and represents the relative size of benefits and costs. It provides an index of the dollar value of benefits that can be expected from a given cost investment, but gives no indication of the scale of investment, which should be considered if alternative control programs are to be compared.
c. *Internal Rate of Return (IRR)*: The **internal rate of return** is the interest rate that would make the total present value of the benefits equal to that of the costs, e.g., to reduce the NPV to zero. The IRR can be easily compared with current interest rates without the necessity of selecting a discount rate. However, the IRR is difficult to calculate and is usually estimated empirically through trial and error.

EXAMPLE 14.2: IS VACCINATION OF CATTLE FOR FOOT AND MOUTH DISEASE IN SOUTH VIETNAM ECONOMICALLY FEASIBLE?

Background: Foot-and-mouth disease (FMD) affects livestock production in a number of ways. Direct impacts are both visible and invisible. The visible damages include draft power loss, milk production loss, abortion, death, and decrease in livestock product value. The invisible losses include reduction in fertility, delay in the sale of animals and livestock products, change in farm structure (resulting from deaths, decreased parturition rate, and delayed sales), and reduced access to markets. Moreover, FMD causes additional expenditures (indirect impacts) in disease control such as vaccination, post-vaccination monitoring, movement restrictions, diagnostic testing, and surveillance. The impact of FMD is especially meaningful to small producers, as it threatens their livelihood and food security. Benefit-cost analysis (BCA) is a commonly used analytical tool that supports the decision-making process in animal disease control. Vaccination is usually implemented twice a year (biannually) in March–April and September–October. The outputs of a BCA would not only foster a review of the national vaccination policy but also provide evidence to encourage farmers' participation in the campaign. Despite its relevance, no BCA for FMD vaccination at the farm level has so far been completed in Vietnam.

Objectives: Truong et al. (2018) conducted a study to analyze the financial impact of foot-and-mouth disease outbreaks in cattle at the farm level and the benefit-cost ratio of the biannual vaccination strategy to prevent and eradicate FMD in cattle in South Vietnam.

Study Design: Cross sectional.

Methods: Two surveys were conducted. An initial questionnaire-based survey captured general information on farm production and management practices from 49 small-scale dairy farms, 15 large-scale dairy farms, and 249 beef farms of Long An and Tay Ninh province. The total number of interviewed farms per district was based on the cattle population density in each district. A stratified sampling strategy for farm selection was used based on the type of cattle production (dairy or beef), with a limit of 10 questionnaires per production type per village. This survey was conducted with the help of veterinary students from Nong Lam University, Ho Chi Minh City, who had received training in interviewing techniques.

In a follow-up survey, financial data on FMD impacts were collected from 129 farms located in 14 villages suspected of experiencing FMD outbreaks (based on owner recognition of clinical signs) using individual semi-structured interviews. In those farms, general data on disease management, control methods, disease impact and all related costs, and specific data on the financial costs associated with FMD were collected using a standardized questionnaire.

The incidence rate of FMD in the study area was estimated from a measured animal level sero-prevalence of 60% among infected herds in the study area, using the formula:

$$\lambda = \frac{-\log(1-p_x)}{x}$$

where λ = the herd incidence rate of FMD, p_x = the measured sero-prevalence in the cattle population, and x = the duration of FMD immunity in cattle (the period during which FMD antibody are detectable after infection). This formula assumes that antibodies to FMD are detectable in cattle for up to 3 years post-infection, and that cattle do not become reinfected following recovery from FMD. The value for λ is used to estimate a number of parameters used in the subsequent partial budget and benefit-cost analyses.

Partial-budget and benefit-cost analyses were carried out as described above and based on the methodological framework described by Rushton (2009). Partial budget analysis was used to estimate the benefits (additional revenue and saved costs) and costs (additional costs and revenue foregone) of using vaccination of one given farm to prevent FMD over a 1-year period according to internationally accepted guidelines for vaccine potency and administration. The net present value of the proposed change in disease control strategy (*alternative scenario*), where FMD vaccination is applied twice a year, was compared to a *status quo* scenario with no FMD vaccination on an individual farm basis for the period of 1 year. The NPV of the alternative scenario was calculated as follows (Table 14.5):

$$\text{Net present value} = (\text{Saved cost} + \text{Additional revenue}) - (\text{Additional cost} + \text{Foregone revenue})$$

The benefit-cost ratio between the alternative scenario and the status quo scenario was also computed on an individual farm basis using following formula:

$$\text{Benefit-cost ratio} = (\text{Saved cost} + \text{Additional revenue}) \div (\text{Additional cost} + \text{Foregone revenue})$$

TABLE 14.5
Partial Budget Analysis Results of the Benefits of FMD Vaccination According to the Different Production Types (Small-Scale Dairy Cattle Farms, Large-Scale Dairy Cattle Farms, and Small-Scale Beef Cattle Farms)[a]

	Small-Scale Dairy Farms	Large-Scale Dairy Cattle Farms	Small-Scale Beef Cattle Farms
Additional cost (kVND)[b]	1120 (459–1922)	3193 (2075–5289)	691 (177–1548)
Foregone revenue (kVND)	3195 (868–6401)	7383 (1542–14,437)	1731 (238–3966)
Saved cost (kVND)	2739 (−17 to 6227)	6466 (−352 to 14,633)	1346 (−8141 to 3667)
Additional revenue (kVND)	17,240 (6523–26,603)	48,548 (33,407–69,647)	2576 (580–5609)
Net present value (kVND)	15,664 (4703–27,202)	44,438 (25,175–65,467)	1499 (−2896 to 5142)

Source: Truong DB, Goutard FL, Bertagnoli S et al. *Front Vet Sci* 2018;5:26. With permission.

[a] Result of Monte Carlo simulation: mean (CI 95%).

[b] kVND: thousands of Vietnam Dong (Vietnamese currency).

The benefit-cost ratio was calculated for three production systems: large-scale and small-scale dairy farm and small-scale beef farm. A sensitivity analysis for benefit-cost of FMD vaccination was performed by changing vaccination cost and market prices of sold cattle and milk.

Results: Applying the above formulae to the data in Table 14.5, the NPV of FMD vaccination in large-scale dairy farms was 2.8 times higher than in small-scale dairy farms and 29.6 times higher than in beef farms. The BCRs of FMD vaccination over 1 year in large-scale dairy farms, small-scale dairy farms, and beef farms were 5.20, 4.63, and 1.62, respectively. Sensitivity analysis revealed that changes in market value had a greater impact on the BCR than changes in vaccination cost for all three production types over the range of values studied.

Conclusions and Significance: This benefit-cost analysis of the current biannual vaccination strategy showed that investment in FMD prevention can be financially profitable, and therefore sustainable, for dairy farmers. For beef cattle, it is less certain that vaccination is profitable. Additional benefit-cost analyses of vaccination strategies at the national level would be required to evaluate and adapt the national strategy to achieve eradication of this disease in Vietnam. As no prior studies on the BCA of FMD vaccination have been conducted in Vietnam, this study's research framework and results are expected to provide a firm basis for further research and awareness programs.

FOLLOW-UP QUESTION 14.2

FMD control through vaccination is a voluntary program in Vietnam, with the exception of government intervention during outbreaks. What considerations should be part of a decision to launch a national FMD eradication program under field conditions? See Answer 14.2 at the end of this chapter.

14.2 DECISION ANALYSIS

In most cases in veterinary practice, the prognosis or economic impact of medical decisions is not certain. The best option, for example, defer treatment, treat empirically, or administer treatment based on the results of diagnostic tests, may not be readily apparent because of the interaction of a

number of variables. If there are multiple possible outcomes of the proposed courses of action, and chance is an important factor in determining which outcome occurs, then **decision analysis** is the approach of choice.

At least four approaches to decision analysis have been described: (1) mathematical equations, (2) payoff matrices, (3) process diagrams or process flow charts, and (4) decision trees. The advantage of decision trees is that they can explicitly depict the chronology of events and can be used to evaluate a sequence of decisions. A decision tree provides the decision-maker with a graphic approach to the decision-making process.

14.2.1 Steps in Building a Decision Tree

Decision tree analysis is a process for analyzing complex choices by the use of decision trees. There are three basic steps in building a decision tree. The first step is to specify the **decision context**, that is, the real-world situation in which a particular decision is to be made. The second step is the development of a **decision model** that includes the management options, the consequences of each option, and how likely and desirable each possible outcome is. The third step is to represent the decision model as a **decision tree** (Figure 14.1), with the consequences of each decision represented by nodes linked by branches.

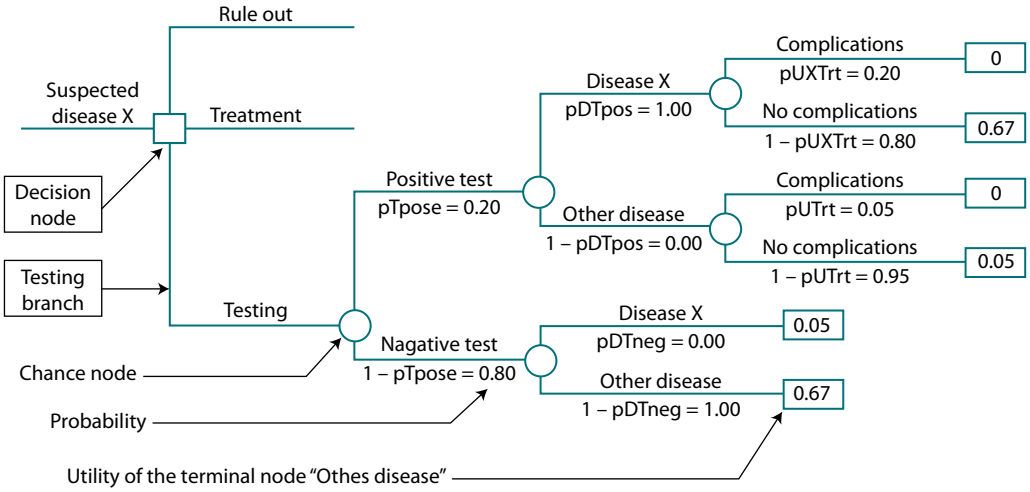

FIGURE 14.1 Diagram of a portion of a decision tree for Disease X to illustrate basic concepts of decision tree analysis. The decision node is designated by a square, chance nodes by circles, and terminal nodes by rectangles. Each of the three branches leading from the decision node represents a different strategic option. The probabilities are located beneath each chance node branch, and utilities are within the terminal nodes. For this example, the prior probability of Disease X has been assumed to be 0.20. Test sensitivity and specificity have been assumed to be 100%, simulating a perfect diagnostic test. Therefore, the probability of a positive test (pTpos) is 0.20, the predictive value of a positive test result (probability of disease given a positive test result, pDTpos) is 1.0, and the predictive value of a negative test result (probability of not having the disease given a negative test result (1 − pDTneg) is 1.0. Baseline values for probabilities and utilities were chosen to approximate average clinical conditions. pUXTrt = probability of complications from treatment of Disease X leading to death. pUTrt = probability of complications leading to death from administering treatment for Disease X to animals suffering from other diseases. By fold back of the tree, the expected utility of a negative test result would be $(0.0 \times 0.05) + (1.0 \times 0.67) = 0.67$. By risk analysis, the probability of death (utility = 0) for the testing branch of this decision tree is $(0.2 \times 1.0 \times 0.2) + (0.2 \times 0.0 \times 0.05) = 0.04$. (From Smith RD. *J Am Vet Med Assoc* 1993;203:1184–1192. With permission.)

14.2.1.1 Nodes

There are three basic types of **nodes**: decision, chance, and terminal. A **decision node** represents a choice between two or more options, such as the decision to test or not to test. **Chance or probability nodes** represent events that are at least partially determined by chance, such as the likelihood that disease is present or that a test result is correct. These probabilities can be assessed from literature, experimental data, or expert opinion. A **terminal node** represents a final outcome with no further significant options or consequences.

14.2.1.2 Utilities

The desirability of a final outcome is expressed as the **utility of a terminal node**. Utility is any measurement that can be used to compare outcomes and determine which outcome is more desirable. The value of each utility is expressed relative to a numerical scale common to all the terminal nodes in the tree. Examples are financial gain (value of the animal minus costs incurred for a particular intervention) or prognosis. The latter is frequently expressed as the probability of short-term survival without sequelae.

14.2.1.3 Variables

Each **variable** in a decision tree must be assigned a baseline value, and the baseline value should approximate the average condition as closely as possible. Two types of variables are found in all decision trees: probability variables and utility. Each of the possible outcomes of a chance node is expressed as a certain probability of occurrence. The sum of the probabilities from each chance node must sum to 100%, or 1.0.

14.2.2 Analysis of the Decision Tree

Once a decision tree is constructed, it can be analyzed by use of techniques for fold back of the tree, sensitivity analysis, and risk profile analysis.

14.2.2.1 Fold Back

In a **fold back**, the **expected utility** for each decision is calculated by adding the values obtained when the utility of each possible outcome of that decision (terminal node) is multiplied by the probability that the outcome will occur. Every fold back starts from some node in the tree, which is referred to as the **root node for the fold back**. In most cases the root node for a fold back is a decision node. The expected utility expresses the average utility of each management option when that option is chosen for a large number of animals. The management option with the highest expected utility is usually the option of choice.

14.2.2.2 Sensitivity Analysis

Sensitivity analysis, which expresses the degree of confidence one can have in a particular decision, is simply a series of fold backs over a range of values for one or more variables. **One-way sensitivity analysis** is used to calculate the changes in expected utility that occur when the value for only one variable is varied. **Two- and three-way sensitivity analysis**, in which two or three values are varied simultaneously, result in a series of **thresholds**, or **break-even points**, at which the expected utility for each decision is equal. The resulting curves are referred to as indifference curves. Threshold values indicate whether a change in a given variable would change the optimal decision (i.e., would result in a different management option being the option of choice) but do not indicate how much would be gained or lost by choosing a given management option.

EXAMPLE 14.3: WHAT IS THE BEST INTRAMAMMARY TREATMENT STRATEGY FOR BOVINE CLINICAL MASTITIS?

Background: Mastitis is the most common and costly health disorder of dairy cows leading to discarded milk, lost production, reduced milk quality, and treatment costs. Clinical mastitis (CM) is often classified according to severity as mild (milk looks abnormal), moderate (milk looks abnormal and the udder or quarter is swollen), or severe (the cow exhibits systemic signs). Short-term clinical and bacteriological outcomes have been reported for cows that received selective treatment of CM based on on-farm culture (OFC) results, but economic outcomes of selective treatment based on OFC had not been reported.

Objectives: Pinzón-Sánchez et al. (2011) used decision tree analysis to evaluate the economic impact of different treatment strategies for the first case of mild or moderate clinical mastitis occurring in early lactation with various scenarios of pathogen distributions and use of on-farm culture.

Study Design: Mathematical (decision tree) model.

Methods: Decisions were ordered to reflect the sequence of decisions made by dairy producers. Economic values and probabilities were derived from the research literature, and expert knowledge (in a few instances where research data were not available). The decision tree (Figure 14.2) included two decision and three probability events. The first decision evaluated use of on-farm culture (two programs using OFC and one not using OFC) and the second decision evaluated treatment strategies (no intramammary antimicrobials or antimicrobials administered for 2, 5, or 8 d). The decision tree included probabilities for the distribution of etiologies (gram-positive, gram-negative, or no growth), bacteriological cure, and recurrence. The baseline distribution of etiologies (scenario A) was based on earlier studies by the author and represented the distribution of pathogens observed on typical large commercial dairy herds located in Wisconsin. Pathogens for scenario A were distributed as 2% *Staphylococcus aureus*, 19% environmental streptococci, 14% coagulase-negative staphylococci (CNS), 24% *Escherichia coli*, 6% *Klebsiella* spp., and 35% no growth. This distribution of culture results was represented in the decision tree as **gram-positive pathogens** (*Staph. aureus*, environmental streptococci, and CNS), **gram-negative pathogens** (*E. coli* and *Klebsiella* spp.), and **no growth**. It was assumed that the diagnosis obtained by using OFC was 100% accurate, although accuracies as low as 80% have been reported. The probabilities of bacteriological cure and of recurrence were based on published research and data collected from four commercial dairy herds, respectively.

The economic consequences of mastitis included costs of diagnosis and initial treatment, additional treatments, labor, discarded milk, milk production losses due to clinical and subclinical mastitis, culling, and transmission of infection to other cows (only for CM caused by *Staphylococcus aureus*). Pathogen-specific estimates for bacteriological cure and milk losses were used. The decision tree had 144 terminal values (utilities) that represented the sum of the partial cash flow (total costs) of each possible outcome. The economically optimal path for several scenarios was determined by comparison of expected monetary values (EMV). Sensitivity analyses were performed using the minimum and maximum values of milk price, cost of farm labor, cost of antimicrobials, and cost of OFC under the baseline prevalence (scenario A). Additional sensitivity analyses were performed by creating two additional scenarios with differing pathogen distributions. Scenario B was characterized by a greater prevalence of CM caused by contagious pathogens (*Staph. aureus*), and scenario C was characterized by a greater prevalence of CM caused by coliforms.

Results: For most scenarios, the optimal economic strategy was to treat CM caused by gram-positive pathogens for 2 d and to avoid antimicrobials for CM cases caused by gram-negative pathogens or when no pathogen was recovered. Use of extended intramammary antimicrobial therapy (5 or 8 d) resulted in the least expected monetary values.

Conclusions and Significance: When CM is treated without knowledge of etiology, it is difficult to justify the routine use of extended-duration therapy for treatment of the first case of CM. Use of OFC is a simple and easy technique that, when used correctly, allows producers

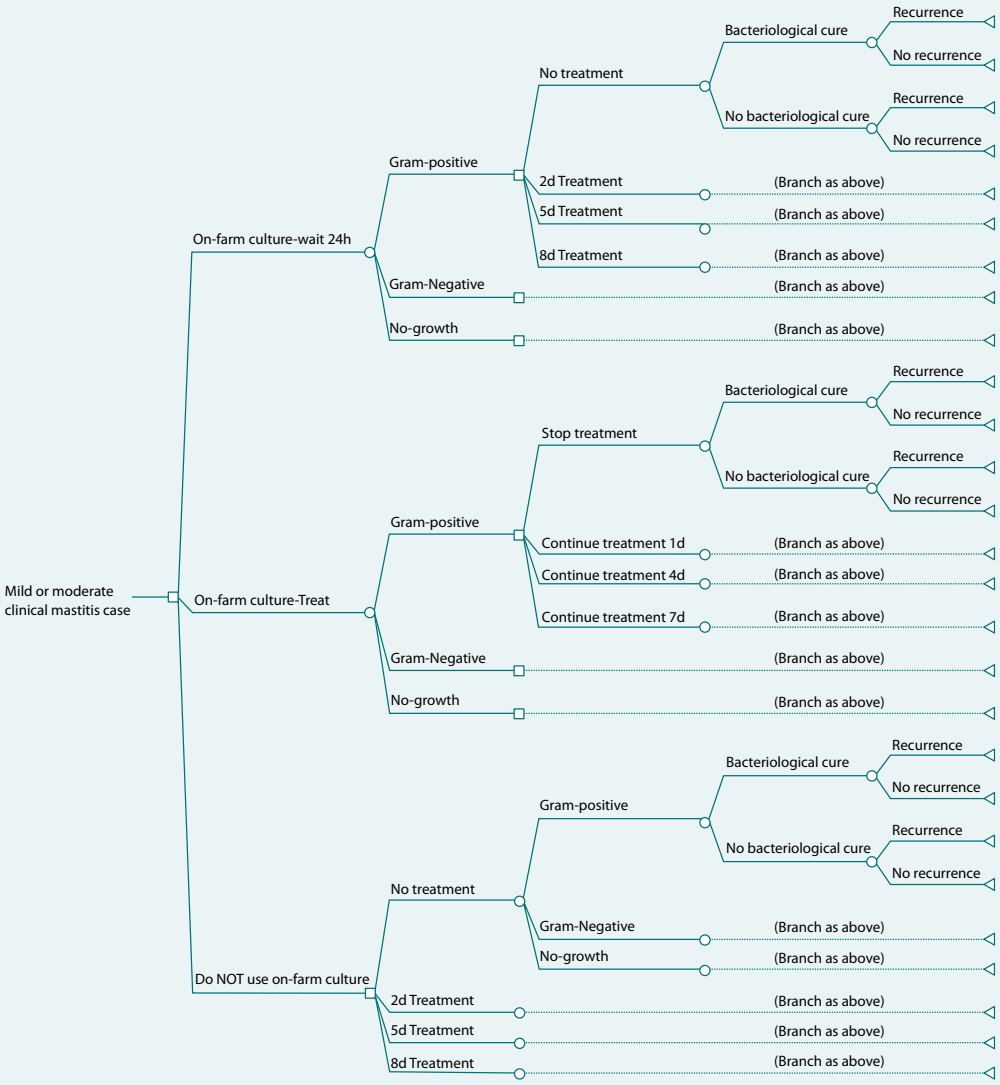

FIGURE 14.2 Simplified structure of the bovine clinical mastitis decision tree. Decision nodes are represented by squares with branches that represent strategies. Probability nodes are represented by circles with branches that represent probability events. Terminal nodes are represented by triangles. (From Pinzón-Sánchez C et al. *J Dairy Sci* 2011;94:1873–1892. With permission.)

to identify the possible pathogen causing CM. Culture-based therapy allowed for the most judicious use of antimicrobials. For most scenarios, the results of the model suggested that the best strategy was to treat mastitis caused by gram-positive pathogens for 2 d and avoid use of antimicrobials for CM caused by gram-negative pathogens or when no pathogen was recovered. Use of extended therapy (5 or 8 d) resulted in the lowest EMV.

Withholding antimicrobial treatment for CM caused by gram-negative pathogens may seem counterintuitive. The generic drug used in the model was assumed to be effective against both gram-negative and gram-positive pathogens, to allow for use for extended-duration therapy, and to require 72 h of milk discard. However, for CM caused by gram-positive bacteria, milk production losses were assumed to persist for the remainder of the lactation, whereas for CM caused by gram-negative bacteria or when no pathogen was recovered (no growth), milk production losses occurred for only 2 months after occurrence of the case. These assumptions may have favored treatment of gram-positive over gram-negative infections in the model.

FOLLOW-UP QUESTION 14.3

Decision tree analysis is an effective method for determining the most economically optimal treatment strategy for commercial dairy herds and is a useful instructional tool to understand the complex interactions affecting the economics of treatment of CM. How could the conclusions drawn from this study be strengthened?

14.2.2.3 Risk Profile Analysis

Fold back of the decision tree does not convey how likely each result is. One may be more concerned with reducing the likelihood of a particular adverse outcome, such as death of the patient, than with obtaining the highest expected utility. **Risk profile analysis** expresses the probability of occurrence of each of the possible outcomes of a particular set of decisions in a decision tree. Starting at the root node, probabilities for each outcome are multiplied consecutively down to each terminal node. The resulting probabilities can be compared to find the set of decisions associated with the lowest risk of an unfavorable outcome. This methodology is a component of scenario tree analysis described below.

14.3 SCENARIO (EVENT) TREES

Scenario trees have been defined as graphical depictions of the biological pathways for hazard introduction into an importing country and have been applied to a variety of animal movement scenarios worldwide (Martin et al., 2007; Peeler et al., 2015). The scenario tree is a structure representing all possible pathways from the starting point (population is infected) to the outcome (infection is detected or not detected) often used to support claims of freedom from disease or infection. The scenario tree structure and components are very similar to those described above for decision tree analysis, but rather than folding back the tree to reveal the best decision option, scenario tree probabilities are multiplied sequentially down the sequence of branches in a forward fashion, ending with the actual likelihood and associated costs of an outcome, usually the introduction of a diseased animal. This is basically an application of risk-profile analysis as described above for the analysis of decision trees. Thus, the combined estimate for the entry and exposure pathways is the likelihood that the pathogen is released into the environment and one or more susceptible individuals exposed (Peeler et al., 2015).

EXAMPLE 14.4: CAN THE COSTS OF *TAENIA SAGINATA* SURVEILLANCE IN CATTLE BE REDUCED BY TARGETED INSPECTION OF HIGH-RISK POPULATIONS?

Background: *Taenia saginata* cysticercus is the larval stage of the zoonotic parasite *Taenia saginata*. Its life-cycle involves both cattle and humans. Humans become infected through ingestion of viable cysticerci in undercooked or raw beef. Cattle are infected through ingestion of feed or water contaminated with human feces containing the eggs of the tapeworm. Despite its low public health impact, the presence of cysts in cattle leads to significant economic losses to the European meat industry and cattle farmers, due to degrading and condemnation of infected carcasses, increased processing costs, and costs relating to meat inspection. The current surveillance system based on post-mortem inspection of carcasses has low sensitivity and leads to considerable economic burden. In addition to examining a large number of animals, the current method also involves incising valued cuts such as the internal and external masseters, leading to an increase in losses. Hence, presently a large amount of money is spent on an inspection method that provides limited public health protection by removing from the food chain only a small fraction of infected carcasses. The accumulation of evidence therefore indicates a situation where a targeted approach to surveillance for *T. saginata* cysticercus in the UK and elsewhere is worth investigating for potential efficiency gains.

Objectives: In order to address both public health concerns and food production efficiency, Chengat Prakashbabu et al. (2018) explored the potential of risk-based and cost-effective meat inspection activities for the detection and control of *T. saginata* cysticercosis in low prevalence settings.

Study Design: Mathematical (scenario tree) model.

Methods: Building on the findings of a study on risk factors for *T. saginata* cysticercus infection in cattle in Great Britain, the authors simulated scenarios using a stochastic scenario tree model (Figure 14.3), where animals are allocated to different risk categories based on their

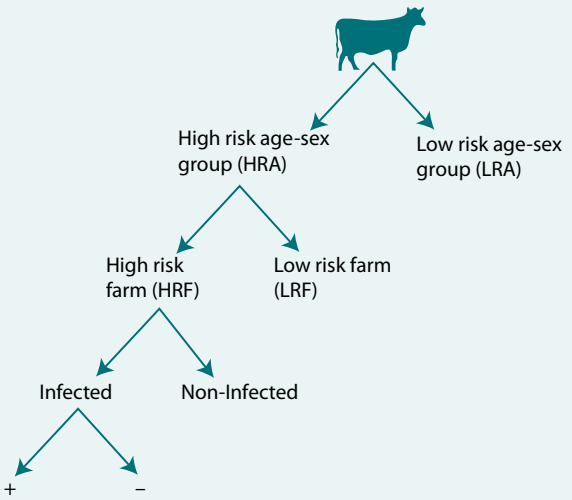

FIGURE 14.3 Scenario tree representation of a risk-based meat inspection system for the detection of *Taenia saginata* cysticercosis in cattle. Animals are divided into different risk categories based on the presence of high-risk farms in their movement history and the age-sex category to which they belong. Each step was assumed to be independent of the others. (From Chengat Prakashbabu B et al. *Parasit Vectors* 2018;11:257. With permission.)

age, sex, and movement history. These animals underwent different types of meat inspection (alternative or current) depending on their risk category. Expert elicitation was conducted to assess feasibility of scenarios and provide data for economic analysis. For each scenario, the technical performance and cost-effectiveness were calculated and results compared. The cost-effectiveness of each scenario was calculated as an incremental cost-effectiveness ratio, using the number of infected carcasses detected as the technical outcome.

Results: Targeting the high-risk population with more incisions into the heart while abandoning incisions into the masseter muscles was found to reduce the total number of inspections and cost while simultaneously increasing the number of infected carcasses found. Under this scenario, the number of inspections required to identify an infected carcass was nearly halved with respect to the baseline scenario and it was cost effective even in the worst-case scenario studied (representing the sensitivity of meat inspection methods).

Conclusions and Significance: The results suggest that, under reasonable assumptions regarding potential improvements to current inspection methods, a more efficient and sensitive meat inspection system could be used on animals categorized according to their risk of harboring *T. saginata* cysticerci at slaughter. Such a system could reduce associated cost to the beef industry and lower microbial contamination of beef products, improving public health outcomes. As an additional benefit, less incision of tissues would not only reduce the cost of inspection but could also reduce the chance of microbial contamination due to a reduction in the handling of muscles and organs at meat inspection.

FOLLOW-UP QUESTION 14.4

What is the next step in the analysis of the proposed risk-based surveillance system?

14.4 STRATEGIES TO REDUCE THE FREQUENCY OF DISEASE

Ultimately, the practitioner must devise a plan for the reduction of disease in the population. This may be accomplished through disease prevention, control (treatment), or eradication. The choice of a particular strategy must be based on an economic evaluation of alternative actions. Most analyses rely on the decision criteria described above. Additional considerations are discussed below.

14.4.1 Disease Prevention

The objective of **disease prevention** is to forestall disease transmission or the occurrence of clinical signs. One way to achieve this is by preventing contact of the host with the agent through isolation, e.g., the removal of a known infected individual(s) from the population, or through quarantine, the confinement of individuals exposed to an infectious agent from other susceptibles. Additionally, animals may be treated prophylactically with antibiotics or immunized to increase their resistance to the agent. Ultimately, disease prevention focuses on the risk factors for disease.

Risk analysis is the formal process for evaluating, managing, determining, and communicating the impact of a risk in a population. It draws on the quantitative measures of risk assessment discussed above and in Chapter 6 and applies them to the prevention of disease at the population level. Risk analysis has become increasingly important in veterinary medicine to identify, evaluate, and mitigate food safety risks, and for decision-making related to international trade and imports (Peeler et al., 2015). It is especially important in mitigating risks associated with the translocation of wildlife for conservation purposes and animal movement into and among zoos.

A risk analysis includes three components: (1) risk assessment, (2) risk management, and (3) risk communication (MacDiarmid and Pharo, 2003). **Risk assessment** deals with identifying and quantifying the actual risks or hazards. **Risk management** is the pragmatic decision-making process aimed at the adoption of actions or policies to mitigate risks. **Risk communication** focuses on communicating risks and options for mitigating risks to the target audience. The objective is to accurately convey the information without over-diagnosing, over-alarming, over-reassuring, or over-planning. It requires a free, unambiguous exchange of information with the affected audience that defines limits of certainty and acknowledges uncertainties involved.

14.4.2 Disease Control

Disease control is aimed at reducing the frequency of disease to a tolerable level. It is usually accomplished through treatment of affected individuals, as during a routine mastitis control program in a dairy. Disease control focuses primarily on the source and mode of transmission of a disease agent.

The level of a disease that is considered tolerable depends on the criteria being used, e.g., whose interests are at stake. Thus, a producer may be striving for certain production indices, the bank manager who loaned money to the producer may be looking at financial returns, and regulatory agencies who inspect the producer's animals or animal products must consider public health risks of the disease.

14.4.3 Disease Eradication

Eradication is the complete elimination of a disease agent from the environment. Eradication may be considered in an individual herd, where the potential for reintroduction of the disease agent can be effectively controlled, or over wide geographic areas.

14.4.3.1 Test and Removal versus Herd Depopulation

Regional or national programs for animal disease eradication may have to decide whether to implement a **test and removal strategy**, where test-positive animals only are removed from the herd, or a **herd depopulation strategy**, where the detection of any test-positive animal results in condemnation of the entire herd. This decision is often based on the sensitivity of the test or test strategy being employed and likelihood of false negative test results (see Chapter 4).

14.4.3.2 Necessary Conditions for Eradication

A number of major livestock diseases and pests have been eradicated from the United States since 1884 (Hagan, 1958). These include contagious bovine pleuropneumonia, glanders in horses, foot and mouth disease, Texas cattle fever, dourine (*Trypanosoma equiperdum* infection) in horses, fowl plague, vesicular exanthema of swine, sheep and goat scabies, hog cholera, and screwworms. A review of these successful eradication programs reveals that the feasibility of eradication depends on meeting one or more of the following conditions:

1. An effective means (diagnostic test) for identification of reservoirs (carriers).
2. An effective method for destruction of the agent in reservoirs (or the reservoirs themselves).
3. A small host range (preferably a single host).
4. A single or limited spectrum of disseminating mechanisms that can be readily manipulated.
5. Acceptability to the industry.

The level of artificially induced herd immunity required to eradicate disease is inversely related to the etiologic agent's intrinsic reproductive number (see Table 12.1). Even highly effective vaccines may be insufficient to eradicate or even control disease if vaccine coverage is inadequate.

REFERENCES

Chengat Prakashbabu B, Marshall LR, Crotta M et al. Risk-based inspection as a cost-effective strategy to reduce human exposure to cysticerci of *Taenia saginata* in low-prevalence settings. *Parasit Vectors* 2018;11:257. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5907745/

Hagan WA. The control and eradication of animal diseases in the United States. *Annu Rev Microbiol* 1958;12:127–144.

Kim I-H, and Jeong JK. Risk factors limiting first service conception rate in dairy cows and their economic impact. *Asian-Australas J Anim Sci* 2019;32:519–526. Full text available at: https://www.ajas.info/journal/view.php?doi=10.5713/ajas.18.0296

MacDiarmid SC, and Pharo HJ. Risk analysis: Assessment, management and communication. In Veterinary services: Organisation, quality assurance and evaluation. Correa Melo E, Gerster F (ed). *Revue Scientifique Et Technique-Office International Des Epizooties*. 2003; 22(2):397–408. Full text available at: http://web.oie.int/boutique/index.php?page=ficprod&id_prec=85&id_produit=520&lang=en&fichrech=1

Martin PAJ, Cameron AR, and Greiner M. Demonstrating freedom from disease using multiple complex data sources 1: A new methodology based on scenario trees. *Prev Vet Med* 2007;79:71–97.

Peeler EJ, Reese RA, and Thrush MA. Animal disease import risk analysis—A review of current methods and practice. *Transbound Emerg Dis* 2015;62:480–490. Full text available at: https://onlinelibrary.wiley.com/doi/full/10.1111/tbed.12180

Pinzón-Sánchez C, Cabrera VE, and Ruegg PL. Decision tree analysis of treatment strategies for mild and moderate cases of clinical mastitis occurring in early lactation. *J Dairy Sci* 2011;94:1873–1892. Full text available at: https://www.journalofdairyscience.org/article/S0022-0302(11)00156-1/fulltext

Rushton J. *The Economics of Animal Health and Production*. Wallingford, UK: CABI; 2009.

Truong DB, Goutard FL, Bertagnoli S et al. Benefit-cost analysis of foot-and-mouth disease vaccination at the farm-level in South Vietnam. *Front Vet Sci* 2018;5:26. Full text available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5834816/

ANSWERS TO FOLLOW-UP QUESTIONS

Answer 14.1: The addition of $567 to the cost of first-service conception failure would increase the attributable risk and population attributable risk in Table 14.4, but would have no effect upon the population attributable fraction, i.e., the proportion of financial loss at the herd level attributable to peri- and postpartum disorders. The reason is that the additional management costs would be shared by all non-conceiving cows, irrespective of risk group (column b in Table 14.3). The population attributable fraction would only increase if the strength of association (between postpartum disorder and failure to conceive) or the proportion of the herd at risk were to increase. Thus, one would have to look elsewhere to reduce the cost of reproductive treatment to achieve conception.

Answer 14.2: The success (or failure) of a formal national FMD eradication program could be influenced by a number of variables whose values should be estimated as accurately as possible before committing to a nationwide FMD eradication program:

- Vaccine coverage, vaccine efficacy (seroconversion rate) and effectiveness (storage, delivery, application)
- Personnel requirements for post-vaccination surveillance and improved reporting of cases
- Need for multiple (cumulative) immunizations over time
- The cost of educational awareness campaigns promoting the benefits of vaccination
- Quantification of the economic losses attributable to the chronic form of FMD through improved record-keeping of cow performance
- Government incentives (subsidies) to promote vaccination
- The cost of cattle movement restrictions in outbreak areas

Answer 14.3: The authors suggest that the biological assumptions of this model could be strengthened by field studies designed to better characterize post-treatment outcomes in dairy cows. Further study

to extend the model to include cows with different days in milk (DIM) or with a previous history of clinical and subclinical mastitis is needed.

Answer 14.4: The next step in the analysis of the proposed risk-based surveillance system should include validation of model predictions with real data. The authors validated key risk probability inputs for targeted inspection of high-risk animals using 6 months of actual slaughter data. The availability of a readily accessible electronic database of all high-risk farms based on movement history and age-sex data will make this risk-based approach feasible in the future. The second component of the system, improved sensitivity of the inspection process through an alternative form of meat inspection that relies on increasing the number of cuts to the bovine heart, has yet to be confirmed.

Glossary

Accuracy: Test accuracy is the proportion of all tests, both positive and negative, that are correct. It is often used to express the overall performance of a diagnostic test. Another name for accuracy is validity.

Adjusted rate: A rate adjusted for subgroup effects by converting their distribution to that of a standard population.

Alpha (Type I) error: Concluding that outcomes are different when, in fact, they are not. Alpha error is analogous to the false-positive result of diagnostic tests (see **beta error** and ***p* value**).

Alternative hypothesis: The alternative to the null hypothesis, i.e., that the observed difference between groups could not have arisen by chance and therefore is real.

Amplifying host: Generally considered to be those intermediate hosts that do not suffer from disease, but in which the number of infectious units increases extensively and provides a source for epidemics in humans or domestic animals.

Apparent prevalence: The prevalence of disease estimated on the basis of diagnostic tests (compare with true prevalence).

Attack rate: The proportion of a defined population affected during a particular outbreak. It is equal to the total number of cases during the outbreak period divided by the number of individuals initially exposed, i.e., those present at the beginning at the outbreak.

Attributable risk (risk difference): The additional incidence of disease attributable to a risk factor itself. It is calculated by subtracting incidence among those not exposed to a risk factor from incidence among exposed individuals.

Basic reproduction number (R_0): The number of secondary cases that one case would produce in a completely susceptible population.

Beta (Type II) error: Concluding that outcomes are not different when, in fact, they are. Beta error is analogous to the false-negative result of diagnostic tests (see **alpha error**).

Between-herd prevalence: The proportion of all herds in which one or more animals possess the condition of interest. Compare with **within-herd prevalence**.

Bias: A preference, inclination, process, or systematic error that inhibits impartial judgment or leads to deviations of results or inferences from the truth.

Blinding: Blinding in research, such as clinical trials, refers to a practice where animal owners and/or clinicians are prevented from knowing certain information about patients under study that may somehow influence them—thereby biasing the results.

Carrier state: A state of infection in which an infected host can communicate the infection in the absence of manifest disease.

Case control (retrospective) study: An epidemiological observational study in which subjects are followed backward in time, from outcomes to possible causes. Cases and noncases are not necessarily members of same population group.

Case definition: The combination of history, physical, or laboratory findings that are characteristic of a particular disease syndrome. It should include all true cases of the disease and exclude similar, but unrelated, conditions. The case definition is the starting point for determining risk, prognosis, or the effectiveness of therapeutic regimens.

Case fatality rate: Number of deaths attributable to a disease during an outbreak divided by the number of cases of that disease during the outbreak period.

Case finding: A strategic form of screening targeted at individuals or groups suspected to be at high risk of infection or disease because of association with known infected or diseased individuals or groups, or through other forms of exposure.

Case report: Detailed presentation of a single case or a handful of cases (<10); may be either cross-sectional or longitudinal.

Case series: Longitudinal study into which subjects are recruited at any stage of their disease. Case series are purely descriptive studies for which the objective is to characterize or summarize the typical clinical course of a particular condition or disease through to some outcome.

Categorical data: See **nominal data**.

Censored observations: Data on patients with incomplete follow-up.

Clinical course of disease: The progression of disease once it has come under medical care (compare with natural history of disease).

Clinical epidemiology: The research discipline concerned with applying epidemiologic methods to questions directly relevant to the practice of medicine at the individual or herd/flock level. Clinical epidemiology focuses on the sorts of questions asked in the practice of medicine. Consequently, the findings have a direct application in medical decision-making. Studies may be observational or experimental.

Clinical significance: The importance in clinical practice of a treatment effect or outcome.

Coefficient of determination (r^2): The square of the Pearson correlation coefficient. A measure of closeness of fit of the data to the linear regression line.

Coefficient of variation: Equal to the standard deviation of a set of measurements divided by their mean. Frequently used to express the precision of clinical measurements; usually expressed as a percentage.

Cohort: A group of individuals that have something in common when they are first assembled, and who are then observed for a period of time to see what happens to them (compare with **survival cohort**).

Cohort (prospective) study: Subjects are followed forward in time, from possible causes to effects. In a concurrent cohort study, the cohort is assembled in the present and followed into the future. In a historical cohort study, the cohort is identified from past records and followed forward from that time up to the present.

Communicable disease: An illness due to a specific infectious agent or its toxic products that arises through transmission of that agent or its products from an infected person, animal, or inanimate reservoir to a susceptible host, either directly or indirectly through an intermediate plant or animal host, vector, or the inanimate environment (Chin, 2000).

Communicable period: The time during which an infectious agent may be transferred directly or indirectly from an infected person to another person, from an infected animal to humans or from an infected person to an animal, including arthropods.

Compliance: The proportion of individuals (or their owners) that adhere to a prescribed treatment regimen. Thus, an efficacious treatment could be ineffective due to poor compliance.

Concordance: Test concordance is the proportion of all test results on which two or more different tests agree. As the number of different tests applied to the same sample increases, the likelihood of agreement on all tests decreases.

Concurrent cohort study: The study group is assembled in the present and followed into their future. This study design usually requires periodic examination of members of the cohort to record new occurrences of the event of interest.

Confidence interval: The theoretical range over which there is a specified probability (usually 95%) of including the true value.

Confounding: When two or more variables found to be associated with an outcome (as disease) are also associated with each other. As a result it is impossible, from a cursory examination, to determine which variable is responsible for the observed outcome. The confounder's association with the outcome may be causal, or it may simply be associated with the true cause.

Congenital transmission: Transmission occurring at, and usually before, birth transovarially, via the placenta, or via the colostrum.

Contagious infection: A transmissible infection that is spread only as the result of an intimate association or contact with infected animals or their excretions or secretions.

Continuous data: See interval data.

Convenience sample: Study participants are self selected.

Correlation coefficient (*r*) or Pearson *r*: See **Pearson product-moment correlation coefficient**.

Covariance: The situation in which the initial values for animals in each experimental group will influence subsequent values. Covariance is of concern in regression analysis where variables, other than the one under consideration, may influence the outcome.

Cross-sectional study: A study in which all observations on a subject are made at essentially one point in time in the course of that subject's illness.

Crude death rate: Number of deaths during an outbreak/mean population during the outbreak period.

Crude rate: An overall rate defined by the formula: (number in entire population with characteristic of interest) ÷ (total number in entire population). Compare with specific rate and adjusted rate.

Cyclical changes: Increases or decreases in rates (such as disease incidence) developing at intervals longer than a year.

Decision tree: A graphical depiction of the process for analyzing complex choices that includes the management options, the consequences of each option, and how likely and desirable each possible outcome is.

Descriptive epidemiology: Descriptive epidemiology endeavors to describe and quantify the distribution of diseases and associated factors in terms of individuals, place, and time. Results are typically expressed as rates, which require numerator (affected individuals) and denominator (population at risk) data.

Diagnostic testing: Use of a test to discriminate animals that have the disease in question from those that have other diseases that compete with the disease of interest in the differential diagnosis. Diagnostic testing begins with diseased individuals.

Dissemination: See **mode of spread**.

Ecological epidemiology: Ecological epidemiology focuses on understanding the important factors that affect transmission of particular disease agents. These factors are frequently referred to as the "agent-host-environment triad."

Effective contacts: Contacts between infected and susceptible animals that result in transmission of a disease agent.

Effectiveness: A measure of how well a treatment works among those to whom it is offered (compare with efficacy).

Efficacy: A measure of how well a treatment works among those who receive it (compare with effectiveness).

Endemic disease: A disease that occurs with predictable regularity in a population unit with only relatively minor fluctuations in its frequency (see **epidemic [epizootic] disease** and sporadic).

Epidemic (epizootic) disease: A disease whose frequency in a population during a given time interval is clearly in excess of its expected frequency, as during an outbreak (compare with endemic and sporadic).

Epidemiology: The study of health and disease in populations. Epidemiology involves (1) the observational study of naturally occurring versus experimentally induced disease, (2) the study of disease in the population versus the individual, and (3) the detection of associations by inferential methods versus the study of pathologic mechanisms.

Epizootiology: Sometimes used in reference to epidemiological studies in animal populations.

Etiologic epidemiology: Etiologic epidemiology is primarily concerned with establishing causal relationships in diseases of undetermined origin. Other terms that have been used to describe this activity are "medical detection" and "shoe-leather" epidemiology.

Evidence-based medicine (EBM): The judicious integration of best research evidence with the patient's values to make decisions about medical care. In ***evidence-based veterinary medicine*** (EBVM), the concept of patient values is most often replaced with the values of the owners, managers, and veterinarians involved in the care of patients.

Experimental study: Epidemiologic study in which the researcher tries to alter the course of events by manipulating the conditions of the experiment. Experimental studies may evaluate the relative merits of various therapeutic, surgical or preventative measures for a particular disease syndrome (compare with observational study).

Explanatory trial: A trial that seeks to assess the **efficacy** of a treatment. Treatment outcomes are measured only in those patients who actually receive it, regardless of where they were originally assigned. Results are typically subjected to a **per-protocol analysis** to assess outcomes.

External validity (generalizability): The degree to which results of a study can be generalized to the population at large from which the sample was drawn, e.g., the **target population**. Compare with internal validity.

Extrinsic incubation period: The period of time between infection of a biological vector and acquisition by the vector of the ability to transmit the agent to another susceptible vertebrate host.

False-negative rate: The likelihood of a negative test result in patients known to have the disease (pT−/D+). It equals (1 – sensitivity).

False-positive rate: The likelihood of a positive test result in patients known to be free of the disease (pT+/D−). It equals (1 – specificity).

Gold standard: The gold standard refers to the means by which one can determine whether a disease is truly present. Its function is that of a quality-control device.

Herd health/preventive medicine: Herd health/preventive medicine endeavors to use epidemiologic information to design optimal disease prevention strategies. Economic considerations, expressed either as cost-effectiveness or cost-benefit, frequently determine which strategy is most effective.

Herd immunity: The proportion of animals in a population that are resistant to infection or disease.

Herd immunity threshold: The proportion of the population that must be protected to achieve eradication.

Herd retest: Herd retest is a modification of serial testing with the exception that test-negative animals, rather than test-positive animals, are retested. The net effect is to ask the herd to prove that it is free of the condition being sought, thereby increasing test sensitivity at the herd level.

Historical cohort study: The study group is assembled from past records and followed into their future, usually up to the present. The term **retrospective cohort** is also used to describe a historical cohort. The term *cohort* is used because every individual has an equal chance of being included in the study, e.g., sampling based on exposure. The term *retrospective* is used because evidence of exposure is based on past records or recall.

Horizontal transmission: Transmission of an infectious agent between contemporaries, or animals of more or less the same generation (see **vertical transmission**).

Iatrogenic: Induced in a patient by a physician's words or actions.

Incidence: The proportion of individuals that develop a condition of interest over a defined period of time. Incidence takes into account new cases only, i.e., cases that have their onset during the time period specified. It is, therefore, a measure of the risk of becoming a case over a defined time period.

Incidence density: A way of expressing incidence where the denominator is not the number of animals at risk for a specific time period, but rather animal time at risk. An incidence of this type is expressed as the number of new cases per total number of animal days or years at risk.

Incorporation bias: Occurs when the diagnostic test being evaluated, or a related test, is also used to support the diagnosis of the disease.

Intention-to-treat (ITT) analysis: Considers the outcome for all subjects entered into a clinical trial, regardless of whether they received the treatment they were actually supposed to receive, i.e., analysis according to treatment assigned rather than treatment received. It is a measure of treatment effectiveness. Compare with per-protocol analysis.

Interaction (effect modification): When one variable in a suspected cause-effect relationship modifies the effect of another.

Internal validity: The extent to which conclusions drawn from a study are correct for the sample of patients being studied. Compare with external validity.

Inter-observer variability: Variation between the results obtained by two or more observers examining the same material or patient.

Interval (continuous) data: Data that are ordered and for which the size of the intervals is known.

Intrinsic incubation period (incubation period): The period of time between infection of the vertebrate host and the appearance of clinical signs.

Intrinsic reproductive number (basic reproductive number): The average number of secondary infections generated by one primary case in a susceptible population; can be used to estimate the level of immunization or other risk reduction strategy required to reduce or eliminate disease.

Investigator bias: In clinical trials, expectations of a response that may influence the subjective interpretation of the results of animal studies and erroneously attribute a response to either the placebo or treatment.

Irregular variation: Reflects random variation in disease occurrence among individuals in a population.

Kappa (k) statistic (Cohen's *k* coefficient): The proportion of potential agreement beyond chance exhibited by two or more tests.

Latency: A state of infection in which an agent is quiescent in a host and, therefore, difficult to detect; implies a potential for activity.

Life table analysis: A method for analyzing the survival of a cohort of patients where the probability of surviving during each time interval is calculated as the ratio of the number of patients surviving to the number at risk of dying during the interval. The chance of surviving to any point in time is obtained by multiplying the probability of surviving during the time interval by the probability of surviving up to the beginning of that interval. The technique can be used to describe other outcomes of disease besides death such as recurrence of tumor, remission duration, rejection of graft, or reinfection (see **survivorship curve**).

Likelihood ratio: A single measure that summarizes a test's performance. It expresses the likelihood that a given test result would occur in a patient with a disease compared to the likelihood of the same result in a patient without that disease. The likelihood ratio for a positive test result is the ratio of the likelihood of a positive result in patients with disease to the likelihood of a positive result in patients without disease (true-positive rate/false-positive rate). The likelihood ratio for a negative test result is the ratio of the likelihood of a negative result in patients with disease to the likelihood of a negative result in patients without the disease (false-negative rate/true-negative rate).

Longitudinal study: Subjects are observed over a period of time, either retrospectively (patient history and medical records) or prospectively (through follow-up).

Management trial: A trial that seeks to determine how **effective** a treatment is among those to whom it is offered, even if the clinician or owner ultimately decides not to follow the original experimental protocol. Results are typically subjected to an **intention-to-treat analysis** to assess outcomes.

Mark-recapture: A technique for estimating total population size (N) from the number sampled (n), based on the proportion of marked animals (M) that are recaptured (m) where $N = n(M/m)$.

Mass screening (surveillance, monitoring): The application of screening tests to large unselected populations. Identification of an affected population may then lead to case finding through testing of each animal in the herd.

Measures of effect: Measures of the association between exposure and disease. Included are relative risk, attributable risk, population attributable risk, and population attributable fraction.

Meta-analysis: A systematic, quantitative method for combining information from multiple studies in order to derive the most meaningful answer to a specific question.

Misclassification bias: A form of **information bias** where the assignment of subjects to groups (such as cases or controls, or exposure status) are erroneous.

Mode of spread: Refers to how a disease agent is spread from one geographic area to another. Synonymous with dissemination.

Mode of transmission: The way(s) in which an etiologic agent is transmitted from affected to susceptible individuals.

Monitoring: See **surveillance**.

Morbidity rates: Direct measures of the commonness of disease in a population. Examples are attack rate, incidence, and prevalence (see **vital statistics**).

Mortality rate: An incidence rate in which the numerator is the number of deaths occurring in a population over a defined period of time. The denominator is the population at risk over that time period.

Moving average: A moving average is a series of data averages centered at each successive measurement point on the time scale.

Multicollinearity: Occurs when two or more predictor variables (covariates) are also highly correlated with each other, independent of the outcome variable.

Natural history of disease: The evolution of disease without medical intervention (compare with clinical course of disease).

Negative correlation: See **correlation coefficient**.

Nominal (categorical) data: Data that can only be placed into categories, without any inherent order. For analytic purposes nominal data is treated as discrete variables.

Nonrandomized controlled clinical trial: Patients are allocated to concurrent comparison groups by means of some nonrandom process (e.g., convenience, clinical judgment, owner preference).

Nosocomial: Hospital-acquired, such as an infection or disease.

Null hypothesis: The hypothesis, or operational assumption, that no difference exists between treatment groups. Observed difference are due to chance.

Objective data: Measurable indices such as temperature, pulse, respiration, results of parasitologic examinations, complete blood counts, radiographs, etc.

Observational study: Epidemiologic study in which the researcher is merely an observer and does not interfere with the natural course of events. Observational studies focus on such things as assessment of risk, cause, or prognosis (compare with experimental study).

Odds ratio: The odds that a case is exposed divided by the odds that a control is exposed to a risk factor. The odds ratio provides a measure of risk for case control studies that is conceptually and mathematically similar to the relative risk obtained in cohort studies, i.e., the stronger the association between exposure and disease, the higher the odds ratio.

One Health: An initiative that focuses on delivering collaborative, multidisciplinary solutions to complex problems at the animal, human, and environmental interface.

Ordinal data: Data in which the order is known (small to large, good to bad, etc.), but the size of the intervals between values is not.

Outbreak: A sudden occurrence of a large number of cases of disease in a short period of time. The actual number of cases that constitutes an epidemic depends on the past history and seriousness of the condition.

Outbreak period: Period of time over which the first and last cases occurred in a population during an outbreak.

***p* value:** The likelihood that an observed result could have arisen by chance alone.

Pandemic: A very large-scale epidemic, usually involving several countries or continents.

Parallel testing: The performance of two or more tests on a patient or herd, usually at the same time. A positive diagnosis requires that only one of the test results be positive. The net effect of parallel testing is to ask the patient to prove that it is healthy.

Patency: A state of infection in which an agent can be recovered or identified from blood or tissues.

Pathogenicity: A measure of an agent's ability to induce disease (see **virulence**).

Pathognomonic: Specifically distinctive or characteristic of a disease or pathologic condition and rarely found in healthy individuals or those afflicted with clinically similar conditions; a sign or symptom on which a diagnosis can be made.

Pearson product-moment correlation coefficient (Pearson *r*): A measure of the strength and direction of a linear (straight line) association between two interval-level variables. The square root of the coefficient of determination, r^2.

Performance bias: Occurs when prior knowledge of which group animals belong to results in differences in care levels, making it difficult or impossible to conclude that a drug or other intervention caused an effect, as opposed to level of care. See also work-up bias.

Period prevalence: Number of cases (old and new) detected over a time period ÷ number of animals examined over the same time period.

Per-protocol analysis: Considers the outcome only for subjects in a clinical trial that actually received an intervention, i.e., analysis according to treatment received rather than treatment assigned. Compare with intention to treat analysis.

Placebo: In clinical trials, an intervention that is indistinguishable from the active treatment, but does not possess its specifically active ingredient.

Point prevalence: Number of cases (old and new) detected at a particular point in time ÷ number of animals examined at the same point in time.

Population at risk: Population group in which an event could occur.

Population attributable fraction: The fraction of disease occurrence in a population that is associated with a particular risk factor. It is estimated by dividing the population attributable risk by the total incidence of disease in the population.

Population attributable risk: A measure of the excess incidence of disease in a population that is associated with the occurrence of a risk factor. It is the product of the attributable risk and the prevalence of the risk factor in a population.

Positive correlation: See **correlation coefficient**.

Post-test probability (or posterior likelihood) of disease: The likelihood that a patient has a disease or condition based on a particular test result. Besides the test result, the post-test probability is influenced by the pretest probability of disease and the sensitivity and specificity of the test being used.

Power of a study: The probability that a trial will find a statistically significant difference when a difference really exists. A powerful study has a higher probability of rejecting the null hypothesis when it should be rejected. Power is analogous to the sensitivity of a diagnostic test and is equal to 1 minus the probability of a beta error.

Predictive value: The probability of a disease, given the results of a test, is called the predictive value of the test. **Positive predictive value** is the probability of disease in an animal with a positive (abnormal) test result. **Negative predictive value** is the probability that an animal does not have the disease when the test result is negative (normal).

Prepatent period: The period of time between infection of the vertebrate host and detectability of an agent in secretions, excretions, blood, or tissues.

Pretest probability (or prior likelihood) of disease: The likelihood that a patient has a disease or condition before a test is run. The pretest probability of disease may be based on a number of parameters including a clinician's experience with similar patients, the prevalence of the

condition in the population from with the individual was drawn, or the post-test probability estimated from one or more previous test results.

Prevalence: The proportion of sampled individuals possessing a condition of interest at a given point in time. It is measured by a single examination of each individual of the group. Prevalence can be likened to a "snapshot" of the population and includes both old and new cases. It is a measure of the risk of being a case at a given moment.

Prevalence survey: Cross-sectional study of a defined population; commonly used in outbreak investigations.

Prognosis: The prediction of the future course of disease following its onset.

Prognostic factors: Conditions that, when present in individuals already known to have disease, are associated with an outcome of the disease.

Randomized controlled clinical trial: Subjects are randomly allocated into treatment and control groups.

Rate: A fraction in which the numerator is included in the denominator.

Ratio: A fraction in which the numerator is not included in the denominator.

Receiver operating characteristic (ROC) curve: A plot of the true-positive rate (sensitivity) on the vertical axis against the false-positive rate (1 – specificity) on the horizontal axis. The ROC curve provides a standard approach to the evaluation of diagnostic test performance.

Relative risk (risk ratio): The ratio of incidence in exposed individuals to incidence in nonexposed individuals. Relative risk is an index of the strength of the association between exposure and disease. If no additional risk is associated with exposure, then both incidences should be equal and the ratio would be equal to 1.

Reliability: A measure of the repeatability or reproducibility of a clinical measurement. Reliability is sometimes referred to as precision.

Reproducibility: Test reproducibility refers to the degree to which repeated tests on the same sample(s) give the same result.

Review bias: Occurs when the results of a test affect the subjective review of the data that establish the diagnosis.

Revised likelihood of a disease: See **post-test probability**.

Risk factors: Factors that are associated with an increased likelihood of an event occurring, such as disease.

Route of infection: The route by which an etiologic agent gains access to the body of a susceptible individual.

Scenario tree: A structure representing all possible pathways from the starting point (population is infected) to the outcome (infection is detected or not detected) often used to support claims of freedom from disease or infection.

Screening: The presumptive identification of unrecognized disease or defect in apparently healthy populations.

Seasonal fluctuations: Regular changes in disease frequency with periods shorter than a year.

Secular trends: Overall long-term rises or declines in disease frequency that occur gradually over long periods of time.

Selection bias: Occurs when the way in which subjects are assigned to study groups influences the results.

Sensitivity: Test sensitivity is defined as the likelihood of a positive test result in individuals known to have the disease or condition being sought. Test sensitivity is sometimes referred to as "operational sensitivity," to distinguish it from "absolute sensitivity," a term used to express the detection limits of an assay.

Serial testing: The retesting of animals that initially tested positive on a different test. All test results must be positive for a positive diagnosis to be made. The net effect is to ask the individual to prove that it is truly affected by the condition being sought.

Sign: An indication of the existence of something; any objective evidence of a disease, i.e., such evidence as is perceptible to the examining physician, as opposed to the subjective sensations (symptoms) of the patient.

Signalment: The systematic description of an individual for purposes of identification (age, breed, sex, identifying marks, etc.).

Spearman's rank correlation coefficient (r), or Spearman rho: The counterpart of the Pearson correlation coefficient for ordinal data.

Specific rate: A rate for a specific subgroup of a population of interest (example: 3–5 year age group). Compare with crude rate and adjusted rate.

Specificity: Test specificity is defined as the likelihood of a negative test result in individuals known to be free of the disease or condition being sought.

Sporadic disease: A disease that occurs rarely and without regularity in a population unit (compare with endemic and epidemic).

Standard deviation: A measure that is used to quantify the amount of variation or dispersion of a set of data values. It is equal to the square root of the **variance**.

Standard error (of the mean; SEM): The **standard deviation of the mean (SEM)**, estimated by dividing the standard deviation of the population by the square root of the number of observations in the sample.

Standard population: A population in which the population characteristics of age, breed, sex, etc. are known and used as a standard. When populations are to be compared, they should have similar components, so usually they are mathematically adjusted to have the same proportions as a standard population.

Statistical significance: A level of confidence in the results of a study based on a predefined *p*-value. Generally refers to *p*-values falling below 0.05, i.e., we are willing to be wrong 5% of the time.

Subjective data: Findings such as general condition, alertness, appetite, bowel movements, urination, evidence of pain, etc., which are based on our own observations and those of the owner.

Surveillance (mass screening, monitoring): The ongoing systematic and continuous collection, analysis, and interpretation of health data for the purpose of monitoring the spatial and temporal patterns of one or more diseases and their associated risk factors.

Survivorship curve: Graphic representation of the number or proportion of a cohort of patients with a particular condition remaining at different points throughout the course of their illness. The technique can be used to describe other outcomes of disease besides death, such as recurrence of tumor, remission duration, rejection of graft, or reinfection (see **life table analysis**).

Sylvatic: Affecting wild animals.

Symptom: Any subjective evidence of disease or of a patient's condition, i.e., such evidence as perceived by the patient; a change in a patient's condition indicative of some bodily or mental state.

True prevalence: The prevalence of disease estimated through use of an appropriate gold standard (compare with apparent prevalence).

Transmissible (communicable) infection: An infection that can be passed from infected to susceptible animals.

Type I error: See **alpha error**.

Type II error: See **beta error**.

Unapparent infection: The presence of infection in a host without recognizable clinical signs or symptoms. Unapparent infections may be identified by laboratory means, including immunologic tests. *Synonyms*—asymptomatic, subclinical, occult infection.

Uncontrolled clinical trial: Clinical trial with no concurrent comparison group.

Validity: The degree to which a measurement reflects the true status of what is being measured. Another name for validity is accuracy.

Variance: The squared deviation of a set of random variables from their mean. Informally, it measures how far a set of (random) numbers are spread out from their average value. The square root of the variance is equal to the **standard deviation**.

Vertical transmission: Transmission of an infectious agent from animals of one generation to animals of the succeeding generation, sometimes transovarially, in utero, or with colostrum (see **horizontal transmission**).

Veterinarian-client-patient relationship: Recognized by the Food and Drug Administration when a veterinarian in a practice (1) has seen the animals to be treated, (2) is familiar with the premises and management system, and (3) has established a tentative diagnosis for the condition to be treated.

Virulence: A measure of an agent's ability to induce severe disease (see **pathogenicity**).

Vital statistics: Rates or population indices that provide indirect evidence of the health status of a population. Examples are birth, fertility, and death rates (see **morbidity rates**).

Within-herd prevalence: The prevalence of a condition of interest among animals within the same herd. Compare with **between-herd prevalence**.

Work-up, detection, or verification bias: When prior knowledge of an animal's test result or treatment status affects the rigor with which the patient is examined. See also **performance bias**.

Index

A

B

C

D

E

N

O

P

Q

R

S

T

U

V

W

Y

Z