Second Edition

Veterinary Clinical Epidemiology

A Problem-Oriented Approach

Ronald D. Smith

1,980-

593 -

Library of Congress Cataloging-in-Publication Data

Smith, Ronald D.
Veterinary Clinical Epidemiology: A Problem-Oriented Approach / by Ronald D. Smith.
2nd ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-8493-2445-9 (alk. paper)
1. Veterinary Clinical Epidemiology. I. Title.
[DNLM: 1. Epidemiologic Methods — veterinary. SF 780.9 S657v 1995]
SF780.9.S62 1995
636.089'44 — dc20
DNLM/DLC
for Library of Congress 94-36939
CIP

International Standard Book Number 0-8493-2445-9
Library of Congress Card Number 94-36939
Printed in the United States of America 1 2 3 4 5 6 7 8 9 0
Printed on acid-free paper

THE AUTHOR

Ronald D. Smith, D.V.M., Ph.D., is Professor of Epidemiology and Preventive Medicine in the College of Veterinary Medicine at the University of Illinois. He received his D.V.M. from Michigan State University in 1967 and his M.S. and Ph.D. 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 staff of the University of Illinois College of Veterinary Medicine in 1974. Dr. Smith's research interests have focused on the 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. These consultancies have focused on the diagnosis, epidemiology, and control of vector-borne blood diseases of animals, and more recently veterinary medical informatics.

He teaches professional and graduate courses on veterinary epidemiology, food hygiene and public health, and medical informatics. Dr. Smith has presented numerous invited papers at international conferences and is principal or coauthor of more than 60 research publications.

PREFACE TO THE FIRST EDITION

Medical knowledge is not static. Approaches to the diagnosis, treatment and prevention of disease change as new medical information is acquired. Much of this information is based on the observation of naturally or spontaneously occurring disease. The science of epidemiology evolved from the need to draw accurate conclusions from the study of health and disease in populations by controlling for bias, confounding and chance. Clinical epidemiology focuses on the application of epidemiologic methods and findings to medical decision-making. Results are usually directly applicable to patient care. Epidemiologic principles are also fundamental to critical interpretation of the medical literature.

This book is not intended to make epidemiologists out of veterinary students, but rather to show how experience with patients can be used to explore issues of importance in the practice of veterinary medicine. The decision to focus on clinical epidemiology in an introductory book for veterinary students was influenced by the following observations: (1) most veterinary graduates go into practice; (2) all practitioners are exposed to epidemiologic data from their patients, scientific meetings and the veterinary literature; and (3) the science of epidemiology plays a significant role in medical decision-making.

The first part of the book focuses on the application of epidemiology in medical decision making at the individual and herd levels. The second part examines the epidemiology of disease in populations and outbreak investigation. Wherever possible, important concepts are illustrated with examples from the veterinary literature. Case studies appear throughout the book. A glossary of epidemiologic terms is also included.

It is the intent of the author that this book serve not only as a teaching resource, but also as a reference manual on the application of epidemiologic methods in veterinary clinical research. Readers' suggestions and contributions will be welcomed.

Ronald D. Smith, D.V.M., Ph.D.
Urbana, Illinois

PREFACE TO THE SECOND EDITION

Since publication of the first edition of this book, the approaches and techniques of clinical epidemiology have become increasingly prominent in the veterinary literature. This second edition includes numerous updates throughout to reflect the increasing recognition of the role of clinical epidemiology as a basic science in clinical research. The chapters on the evaluation and use of diagnostic tests include expanded sections on likelihood ratios and ROC curves. The chapter on evaluating the cost of disease includes an expanded section on decision analysis. Many of the examples throughout the book have been updated with more recent examples from the veterinary literature.

During the revision process I have tried to maintain the basic focus of the book, e.g., the application of epidemiologic principles and techniques to problems regularly faced by veterinary practitioners. It is hoped that the book will help anyone working in the field of animal health to critically evaluate their own experiences and those of others, as reported in the medical literature and other forums.

Ronald D. Smith, D.V.M., Ph.D.
Urbana, Illinois

ACKNOWLEDGMENTS

I want to acknowledge the encouragement and support of Dr. George T. Woods, Professor Emeritus of Epidemiology and Preventive Medicine, College of Veterinary Medicine, University of Illinois, who helped me recognize and pursue my true interests. It was he who sent me into the classroom, thereby planting the seed that led to this book.

I am indebted to the numerous veterinary students whose questions, critiques, and suggestions over the years have helped make the textual material more relevant and intelligible. My colleagues, Drs. Laurie Hungerford, Ron Weigel and Uriel Kitron, have also made many helpful suggestions to ensure the accuracy of the concepts and methods included in the text. Drs. M.D. Salman, G.M. Allen, and R. Ruble provided insightful reviews of the first edition and suggested ways to improve the second edition.

I must also recognize the contributions of the many fine veterinary researchers whose works are cited profusely throughout the text. I especially want to recognize the fruitful exchanges with Drs. Michael Thrusfield and Paul Pion, which prompted a revision of the sections on multiple test strategies and clinical trials.

Finally, I want to thank Mr. Paul Petralia, Life Science Editor, Ms. Heather Grattan, Project Editor, and the rest of the staff at CRC Press, for their guidance and patience during the editorial process.

The task of preparing the second edition of this book was made easier by the continued understanding and support of my family: Lupe, Ronald, and Veronica.

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

I. DEFINITIONS

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

A. "*...the study of the distribution and determinants of disease frequency in man.*" (MacMahon and Pugh, 1970).

B. "*The study of the patterns of disease...*" (Halpin, 1975).

C. "*...the study of the health status of populations...*" (Schwabe et al, 1977).

D. "*...the research discipline concerned with the distribution and determinants of disease in populations.*" (Fletcher et al, 1982).

E. "*...Epidemiology is nothing more than ecology with a medical and mathematical flavor.*" (Norman D. Levine, 1990, personal communication).

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 (Dohoo et al, 1994).

Epidemiology is not limited to the study of disease; it may also be used to determine what keeps a population healthy. Epidemiology may thus be considered as the study of health and disease in populations. This definition alone does not appear to provide sufficient grounds for creating a separate discipline. After all, laboratory researchers study health and 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.

Table 1.1 Clinical issues and questions in the practice of medicine

Normality/Abnormality	What are the limits of normality? What abnormalities are associated with having a disease?
Diagnosis	How accurate are the diagnostic tests or strategies used to find a disease?
Frequency	How often does a disease occur? How common are each of the findings that occur in a disease?
Risk/Prevention	What factors are associated with an increased or decreased likelihood of contracting disease?
Prognosis	What are the consequences of having a disease? What factors are associated with an increased or decreased likelihood of recovering from disease?
Treatment	How effective is a treatment, and how does it change the future course of a disease?
Cause	What conditions result in disease?

From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology - The Essentials*, first edition, Introduction. Copyright 1982, The Williams & Wilkins Company. With permission.

II. EPIDEMIOLOGIC APPROACHES

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.

A. 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 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 Veterinary Medical Data Base (VMDB) is an example of a descriptive, data-gathering technique. Other examples are the National Animal Health Monitoring System (NAHMS), the Market Cattle Testing Program and the Statistical Reports of the Food Safety Inspection Service. Results are expressed as descriptive statistics.

Analytic epidemiology goes beyond the purely descriptive process to draw statistical inferences about disease occurrence and possible causal associations. Techniques employed include risk factor analysis, life table analysis, mathematical modeling, multiple regression and a variety of statistical tests of significance.

B. 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.* Traditionally, ecological epidemiology has focused on the life cycle, or natural history, of disease. 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. Recent advances in molecular biology (monoclonal antibodies, restriction mapping and DNA probes) and in computer science (computer simulation) are contributing to our understanding of the dynamics of disease transmission.

C. ETIOLOGIC EPIDEMIOLOGY

Etiologic epidemiology is primarily concerned with establishing causal relationships for diseases of undetermined origin. Other terms that have been used to describe this activity are "medical detection" and "shoe leather" 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.

D. 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. Economic considerations, expressed either as cost-effectiveness or cost-benefit, frequently determine 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 deal effectively with producers.

E. CLINICAL EPIDEMIOLOGY

Clinical epidemiology focuses on the sorts of questions asked in the practice of medicine (Table 1.1). Consequently, the findings have a direct application in medical decision-making. 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. These studies focus on such things as assessment of risk, cause or prognosis. 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.

III. APPLICATIONS OF EPIDEMIOLOGY IN VETERINARY PRACTICE

A. MEDICAL DECISION-MAKING

Although practitioners rely on continuing education and the medical literature to keep abreast of advances in the field, one's own patients represent an important source of medical information. Most, if not all, practice experience represents clinical epidemiologic data. For

example, a typical practitioner sees many patients over time and keeps records of varying complexity on each of them. In addition to owner and billing information, a patient record includes the age, breed, sex and medical history (clinical and laboratory findings, diagnosis, treatment, outcome) of each. Medical records may be organized in a variety of ways, particularly with the advent of new practice-oriented computer software. The cumulative information contained in a patient data base can help practitioners evaluate and improve their decision-making procedures. The astute observations of a practitioner may provide important information about a disease.

B. CLINICAL RESEARCH

Clinical epidemiologic findings complement laboratory studies of experimentally induced disease in exploring causal relationships in disease. Whereas laboratory studies provide the biological plausibility, epidemiologic studies must be used to determine whether hypothesized mechanisms are important in the field. Some clinical issues cannot be approached in the laboratory. For example, the effectiveness of treatments must be measured in clinical scenarios. Because the data come from actual cases, the findings should be representative of what would occur in one's own patients. Other clinical issues that are difficult to approach experimentally are risk assessment, cause of diseases of multiple or uncertain etiology, and disease prognosis with and without treatment. Clinical epidemiology also provides a means to study rare conditions or complications of disease that would be difficult to induce experimentally.

Practitioners should also be aware of the limitations of clinical research findings. Bias and confounding from the imprecision of case definitions, the difficulty of establishing representative comparison groups, loss of subjects to follow up, and chance can lead to erroneous conclusions.

C. MEDICAL CONTROVERSY

Medicine, like all fields of science, operates under a system whereby hypotheses and practices are continually being challenged and updated by the collective experience of researchers and practitioners throughout the world. New treatments replace old ones, new diseases are "discovered," and disease mechanisms are finally understood. Many medical procedures are on uncertain ground, sure to be replaced over time.

The medical literature is a forum where our current knowledge is continually tested and updated. The reports themselves are subject to bias, methodological errors and invalid assumptions. Consequently, practitioners must continually monitor and critically evaluate the literature to stay abreast of new developments and determine what medical claims are worthy of consideration. Epidemiology provides the tools for critical evaluation of medical claims.

EXAMPLE: A recurring controversy is the extra-label use of drugs in veterinary practice. Food and Drug Administration guidelines state that "*The use or intended use of new animal drugs in treating food-producing animals in any manner other than in accord with the approved labeling causes the drugs to be adulterated under the Federal Food, Drug, and Cosmetic Act*" (AVMA, 1984). Recognizing the need for veterinarians to make decisions on the appropriateness of extra-label use of drugs in food-producing animals, regulatory action would not ordinarily be considered when the health of food-producing animals is immediately threatened and suffering or death would result from failure to treat the affected animals. In addition, all of the following criteria must be met and precautions observed:

- "A careful medical diagnosis is made by an attending veterinarian within the context of a valid *veterinarian-client-patient relationship*;
- A determination is made that (1) there is no marketed drug specifically labeled to treat the condition diagnosed, or (2) drug therapy at the dosage recommended by the labeling has

been found clinically ineffective in the animals to be treated;
- Procedures are instituted to assure that identity of the treated animals is carefully maintained; and
- A significantly extended time period is assigned for drug withdrawl before marketing meat, milk, or eggs; steps are taken to assure that the assigned time frames are met, and no illegal residues occur."

These guidelines imply that individual practitioners must continually evaluate the efficacy and effectiveness of current therapies and make decisions on drug withdrawal times. Clinicians are confronted frequently with similar information about other cause-and-effect relationships that affect their approach to diagnosis, treatment and prevention of disease.

IV. 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 review critically and extract useful information from the medical literature.

A. DEVELOPMENT OF MEDICAL DECISION-MAKING SKILLS

One of the major problems faced by our generation is learning to deal with uncertainty and making decisions in the face of inadequate, incomplete or equivocal data (Gordis, 1980). Nowhere is this more so than in medicine. 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 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 frequency distribution should help you choose and interpret diagnostic tests and make clinical decisions. A practical problem resulting from the frequency distribution is that of "case definition," the starting point for determining the effectiveness of new therapeutic regimens.

EXAMPLE: Two properties of diagnostic tests that affect their performance are sensitivity and specificity. Sensitivity data frequently are not recognized as such when used to describe clinical findings in patients. Table 1.2 summarizes pathologic findings in 100 dogs that succumbed to *Ehrlichia canis* infection. The frequency of gross hemorrhage ranged from 84% in the heart to 4% in the meninges. Which provides better criteria for ruling out canine ehrlichiosis: the absence of cardiac hemorrhage or absence of meningeal hemorrhages?

B. LEARN EPIDEMIOLOGIC METHODOLOGY AND HOW TO ANALYZE AND PRESENT DATA

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. Graphic analysis of data can help clarify relationships and trends.

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 govern-

Table 1.2 Lesions of canine ehrlichiosis based on necropsy and histopathologic examination of 100 dogs dying or killed in extremis

Pathologic Change	*Percentage of Dogs*
Hypoplasia of bone marrow	100*
Plasmacytosis of kidney	93
Centrilobular degeneration of liver	90
Excessive plasmacytosis of lymph nodes	86
Gross hemorrhage of heart	84
Microscopic hemorrhage of heart	70
Hemorrhagic or enlarged lymph nodes	59
Edema of limbs	55
Gross and microscopic hemorrhage of stomach	53
Gross and microscopic hemorrhage of small intestine	52
Gross hemorrhage of urinary bladder	51
Microscopic hemorrhage of urinary bladder	51
Plasmacytosis in retina	43
Gross hemorrhage of kidney	42
Microscopic hemorrhage of kidney	31
Hemorrhagic or enlarged tonsils	24
Emaciated	22
Plasmacytosis, portal triads of liver	18
Microscopic hemorrhage of testicle	18
Nonsuppurative encephalitis	16
Acute centrilobular necrosis of liver	16
Microscopic hemorrhage in eye	13
Plasmacytosis of urinary bladder	12
Gross hemorrhage of testicle	12
Microscopic hemorrhage in meninges	11
Plasmacytosis of third eyelid	10
Gross and microscopic hemorrhage of esophagus	9
Gross hemorrhage in eye	5
Gross hemorrhage in meninges	4
Plasmacytosis of testicle	4
Icterus	3
Microscopic hemorrhage in brain	2

*Nineteen of 19 submitted.

From Hildebrandt, P.K., Huxsoll, D.L., Walker, J.S., Nims, R.M., Taylor, R., and Andrews, M. 1973. Pathology of canine ehrlichiosis (tropical canine pancytopenia). *Am. J. Vet. Res.* 34:1309-1320. With permission.

mental disease control programs is frequently dictated by statistical considerations. Private practitioners may be asked to participate in state and federal regulatory efforts and must understand their scientific basis. Accredited veterinarians are authorized to test animals for brucellosis, tuberculosis and pseudorabies, and to sign health certificates for interstate movement.

EXAMPLE: Industry literature (Straw, 1985) states that a sample of 30 animals can be considered to be representative of an entire swine herd. If you examine 30 carcasses, what is the chance of failing to detect a disease affecting 10% of the herd? A similar problem, and its solution, can be found in Chapter 9 - Statistical Significance (Sample Size).

C. LEARN TO READ THE MEDICAL LITERATURE CRITICALLY

Veterinary journals play an important role in keeping practitioners abreast of current medical knowledge. Examples are reports of new and emerging diseases, risk factors for disease and injury, and prognosis with or without medical intervention. A variety of study designs are used in clinical research (Table 1.3). The usefulness of this information ultimately depends on the adequacy of the study design and the analysis and interpretation of the data.

A variety of study designs are used in clinical research. The poorest designs are so prone to problems of chance, bias and confounding factors that the validity of their conclusions is marginal.

The strength of clinical research designs varies considerably. Each has inherent strengths and weaknesses (Table 1.4). The poorest designs are so prone to problems of chance, bias and confounding factors that the validity of their conclusions is marginal (Dohoo and Waltner-Toews, 1985a-c). 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.

EXAMPLE: Published literature was examined to determine the study designs used and clinical issues examined in a typical veterinary practice journal and to discover ways to improve the effectiveness of these studies (Smith, 1988). A total of 146 articles appearing in 11 of 12 issues of the *Journal of the American Veterinary Medical Association*, volume 189, covering July to December, 1986, were reviewed. Classification keys were used to identify one of nine possible study designs and seven possible clinical issues (Tables 1.1 and 1.3).

Of the 146 articles, which were contributed by 139 different first authors, a total of 153 study design/clinical issue combinations were identified. Only ten (7%) study designs dealt with experimentally induced disease. The remaining 143 (93%) dealt with spontaneously occurring conditions and fell within the discipline of clinical epidemiology (Figure 1.1). Case reports, in which ten or fewer individuals were studied, accounted for 58% of all study designs. They were followed in frequency by prevalence surveys (11%), uncontrolled clinical trials (9%) and case series (8%). Case control, cohort and controlled clinical trials accounted for not more than 2% each of study designs.

Among clinical issues (Figure 1.2), cause was most frequent (44% of all clinical issues), followed by treatment (24%) and frequency (8%). Normality/abnormality, risk and risk prevention, and diagnostic test evaluation occurred with equal frequency (7% each), while prognosis (2%) was the least commonly examined of the clinical issues. Statistical analyses were employed in 32 (22%) of 146 articles.

It was concluded that 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.

Table 1.3 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 ten individuals 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	Go to 4
	Longitudinal - Subjects followed prospectively over a period of time; groups may be formed in the past (from records) or in present	Go to 6
4.	Comparison group absent	Case series
	Comparison group present	Go to 5
5.	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
	Cases and noncases ascertained from a single examination of a defined population	Prevalence survey
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

From Smith, R.D. 1988. Veterinary clinical research: a survey of study designs and clinical issues appearing in a practice journal. *Journal of Veterinary Medical Education* 15(1):2-7. With permission.

Table 1.4 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	Evaluation of diagnostic tests; sensitivity of diagnostic tests
Prevalence survey	Temporal relationships; measures prevalence, not incidence	Incrimination of risk or causal factors; outbreak investigation
Case control	Temporal relationships; bias in selection of comparison group	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 or without 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

Source of data: Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology - The Essentials*, first edition, Introduction.

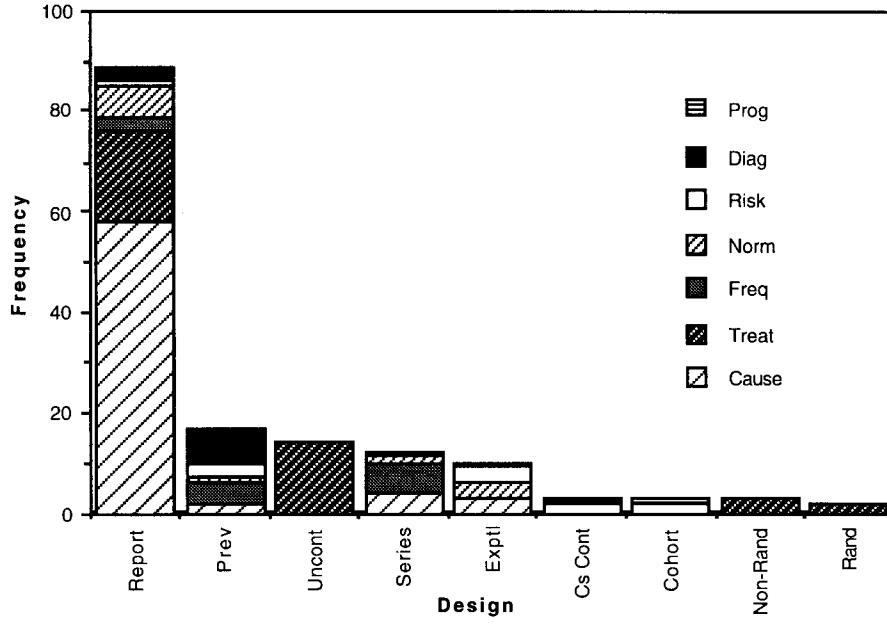


Figure 1.1 Frequency distribution of study designs appearing in 146 articles, subdivided by clinical issue examined. (From Smith, R.D. 1988. Veterinary clinical research: a survey of study designs and clinical issues appearing in a practice journal. *Journal of Veterinary Medical Education* 15[1]:2-7. With permission.)

Legend for study designs: Report = case report; Prev = prevalence survey; Uncont = uncontrolled clinical trial; Series = case series; Exptl = experimental study; Cs Cont = case control; Cohort = cohort study; Non-Rand = non-randomized controlled clinical trial; Rand = randomized controlled clinical trial.

Legend for clinical issues: Prog = prognosis; Diag = diagnostic test; Risk = risk and risk prevention; Norm = normality/abnormality; Freq = frequency; Treat = treatment; Cause = cause.

V. SUMMARY

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.

Over the years a number of approaches and associated methodologies have emerged. Descriptive epidemiology attempts to describe and quantify the distribution of diseases and associated factors in a population or defined geographic region. Ecological epidemiology focuses on understanding the important factors that affect transmission of particular disease agents and produce disease. These factors are frequently referred to as the "host, agent, and environment triad." Etiologic epidemiology is primarily concerned with establishing causal relationships (risk factors) in diseases of undetermined cause. Herd health/preventive medicine endeavors to use information from any or all of the previously mentioned sources to design optimal preventive strategies. Clinical epidemiology is the application of epidemiologic principles and methods to problems encountered in clinical medicine. It focuses on the substance of epidemiologic studies and their practical application to clinical settings.

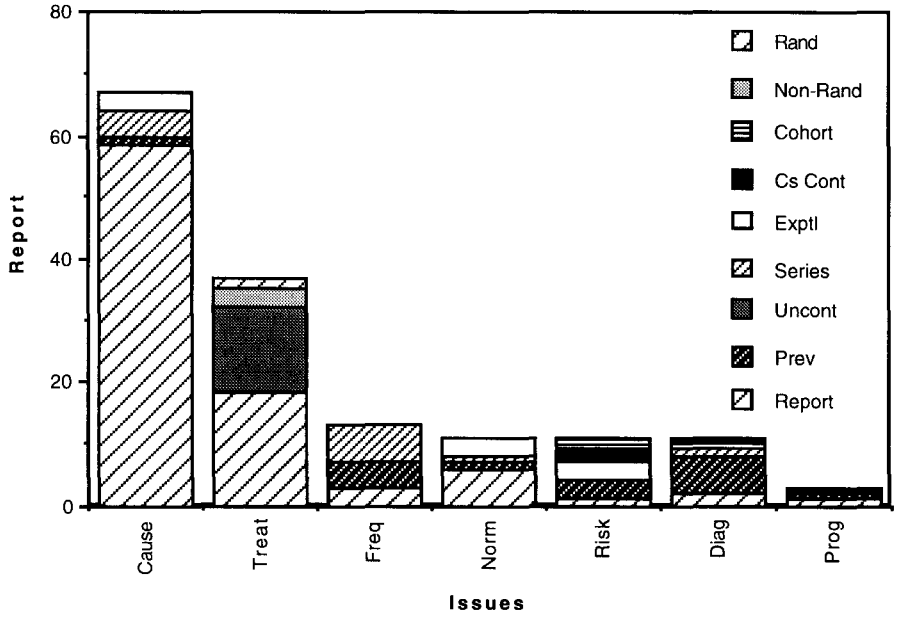


Figure 1.2 Frequency distribution of clinical issues appearing in 146 articles, subdivided by study design employed. (From Smith, R.D. 1988. Veterinary clinical research: a survey of study designs and clinical issues appearing in a practice journal. *Journal of Veterinary Medical Education* 15[1]:2-7. With permission.)

Legend for clinical issues: Prog = prognosis; Diag = diagnostic test; Risk = risk and risk prevention; Norm = normality/abnormality; Freq = frequency; Treat = treatment; Cause = cause.

Legend for study designs: Report = case report; Prev = prevalence survey; Uncont = uncontrolled clinical trial; Series = case series; Exptl = experimental study; Cs Cont = case control; Cohort = cohort study; Non-Rand = non-randomized controlled clinical trial; Rand = randomized controlled clinical trial.

The tools of epidemiology include a variety of techniques for collecting, analyzing and interpreting data. They enable the practitioner to draw accurate conclusions about populations by controlling for bias and random error. The variable manifestations of disease in a population contribute to the uncertainties of medical decision-making. Knowledge of the probabilities or risks associated with a particular cause, outcome or treatment are fundamental to medical decision-making. Because journals play such an important role in the communication of medical information to practitioners, students must learn how to read modern medical journals critically. Much of this information is gathered by straightforward epidemiologic methods including risk assessment, cohort, and case control studies. Basic epidemiologic knowledge is useful for understanding the current medical literature and for interpreting the often conflicting results of clinical studies.

Chapter 2

DEFINING THE LIMITS OF NORMALITY

I. 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 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. First, subjective data is collected, such as alertness, attitude, evidence of pain, etc. These data are based on our own observations and those of the owner. Objective data is collected also; indices include temperature, pulse, respiration, results of parasitologic examinations, complete blood counts, radiographs, etc. This data is 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 ruleout of other possible diagnoses, a treatment or client education (Sandlow et al, 1974).

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 data bases, 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.

II. PROPERTIES OF CLINICAL MEASUREMENTS

Practitioners are continually collecting, categorizing and quantitating 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 popu-

lation values. In this section we examine the factors that influence the confidence we place in clinical measurements.

A. SIGNS AND SYMPTOMS: OBJECTIVE VERSUS SUBJECTIVE DATA

The following are definitions from *Dorland's Illustrated Medical Dictionary* (1981).

- A sign is "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."

- A symptom is "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."

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.

EXAMPLE: Behavioral characteristics are an example of subjective data used to describe animals. Investigators (Hart and Miller, 1985) sought to develop breed behavioral profiles based on 13 traits (Table 2.1) as a guide for potential pet owners. In order to obtain profiles that were quantitative and free of personal biases, they surveyed 48 small-animal veterinarians and 48 obedience judges, randomly selected from directories so as to represent equally men and women, and eastern, central and western regions of the United States. The authors concluded that it is possible to obtain quantitative data that reflect objectively the consensus of authorities about differences in behavior among breeds of dogs. Some behavioral traits discriminated between breeds better than others. The authors attributed this ranking in part to early training and environment.

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

Table 2.1 Behavioral characteristics used as a basis for constructing behavioral profiles of 56 dog breeds (ranked in order of decreasing reliability based on the magnitude of the F ratio)

Behavioral Characteristic	*F ratio**
1. Excitability	9.6
2. General activity	9.5
3. Tendency to snap at children	7.2
4. Excessive barking	6.9
5. Playfulness	6.7
6. Obedience trainability	6.6
7. Watchdog barking	5.1
8. Aggression to dogs	5.0
9. Dominance over owner	4.3
10. Territorial defense	4.1
11. Affection demand	3.6
12. Destructiveness	2.6
13. Housebreaking ease	1.8

* $P<0.005$; see Chapter 9 (Significance) for a more complete discussion of P values.

From Hart, B.L. and Miller, M.F. 1985. Behavioral profiles of dog breeds. *J.A.V.M.A.* 186:1175-1180. With permission.

Data that are ordered and for which the size of the intervals are known are called *interval.* 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.

If there is a mathematically meaningful zero point, the scale may be referred to as *ratio* rather than interval. The Kelvin temperature scale is a ratio scale because $0^{o}K$ indicates an absence of kinetic energy. In contrast, $0^{o}C$ still has measurable heat. Technically, packed cell volume, blood cell counts, and serum biochemical parameters are all ratio-level variables. Since ratio-level variables are treated the same as interval-level variables for statistical purposes, they will be considered as interval data throughout this text.

It is not uncommon for interval-level information 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 wks, 2 wks-2 mos, 2-6 mos, 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 risk or prognostic factors for a specific condition may be impossible without precise age and weight data. Therefore, if time and other limitations permit, information should be recorded at the same level as it was measured.

An example of the differences among nominal, ordinal and interval-level variables is outlined in Table 2.2, which summarizes the clinical assessment of canine and feline anemia.

Table 2.2 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	Onset (acute or chronic), color of mucous membranes, color of stool, degree of lethargy, depression, weakness
Interval	Age, cardiac and respiratory rates, packed cell volume, complete blood count, frequency distribution of erythrocyte morphologic types, total plasma protein

Source of data: Straus, J.H. 1982. Anemia. In, W.R. Fenner (ed), *Quick Reference to Veterinary Medicine.* J.B. Lippincott Co., Philadelphia, pp. 383-398.

C. 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 Tumours 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.

The uniformity of clinical staging among practitioners varies, depending in large part on the subjectivity of the criteria used. For example, contrast the relatively rigid TNM criteria for classification of canine prostate tumors (Turrel, 1987) (Table 2.3) with criteria for depression scores used to estimate acid-base status in diarrheic calves (Kasari and Naylor, 1985) (Table 2.4). 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.

D. 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 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 auscultory findings versus necropsy, or the reproducibility

Table 2.3 Clinical stages of tumors of the canine prostate gland

T	Primary tumor
	T0 = no evidence of tumor
	T1 = intracapsular tumor, surrounded by normal gland
	T2 = diffuse intracapsular tumor
	T3 = tumor extending beyond the capsule
	T4 = tumor fixed or invading neighboring structures
N	RLN*
	N0 = no evidence of RLN involvement
	N1 = RLN involved
	N2 = RLN and juxta-RLN involved
M	Distant metastasis
	M0 = no evidence of distant metastasis
	M1 = distant metastasis detected

*Abbreviation: RLN = regional lymph nodes.
RLN include external and internal iliac nodes; juxta-RLN include lumbar nodes; b = bony involvement.

From Turrel, J.M. 1987. Intraoperative radiotherapy of carcinoma of the prostate gland in ten dogs. *J.A.V.M.A.* 190:48-52. With permission.

of descriptions of the lung sounds reported by different clinicians. The validity of radiographic or serologic diagnoses of occult heartworm disease in client-owned dogs is usually measured by response to therapy, because a truly valid diagnosis can only be made at necropsy.

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.

E. 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.

1. Measurement Variation

Measurement variation may be due to the performance of the instruments being used, the observers themselves or both. 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.

Table 2.4 Clinical signs used to derive depression scores for dehydrated diarrheic calves

Variable	*Method of Assessment*	*Depression Score*	*Interpretation*
Enophthalmos	Visual	0	None
		1	Slight separation of globe and nictitating membrane from eyelids
		2	Marked separation of globe and nictitating membrane from eyelids
Warmth of oral cavity	Fingers in contact with mucosa of hard/soft palate	0	Warm
		1	Cool
		2	Cold
Warmth of extremities	Hand clasped around fetlock	0	Warm
		1	Cool
		2	Cold
Suckle reflex	Index finger over tongue	0	Strong coordinated suckle
		1	Weak coordinated suckle
		2	Chewing motion of jaw
		3	Absent
Menace reflex	Rapid hand movement toward eye	0	Strong, instantaneous eyelid closure
		1	Delayed, slow eyelid closure
		2	Absent
Tactile response	Skin pinched over lumbar area	0	Skin twitch, head movement toward flank
		1	Skin twitch, no head movement toward flank
		2	No skin twitch or head movement toward flank
Ability to stand	Prod thorax with pen	0	Ability to stand
		2	Inability to stand

From Kasari, T.R. and Naylor, J.M. 1985. Clinical evaluation of sodium bicarbonate, sodium L-lactate, and sodium acetate for the treatment of acidosis in diarrheic calves. *J.A.V.M.A.* 187:392-397. With permission.

EXAMPLE: The Schirmer tear test is used routinely in evaluating the adequacy of tear production in animals that have signs of keratoconjunctivitis. Hawkins and Murphy (1986) evaluated the clinical importance of observed discrepancies in the absorptive capacity of tear test strips. Major inconsistencies in the ability of test strips to absorb water were found within one lot of tear strips from a single manufacturer. The variability in the tear strips examined could influence the clinical diagnosis of keratoconjunctivitis sicca and the subsequent interpretation of response to treatment, as well as the interpretation and repeatability of research data.

2. Biological Variation

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 microfilarial counts or body temperature. In others it varies with each patient.

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: Nutrient foramina are common findings in skeletal radiography. Location and radiographic appearance of these foramina are usually uniform and bilaterally symmetrical. Foramina that appear in unusual locations may be misdiagnosed as fractures. Losonsky and Kneller (1988) examined bilateral metacarpophalangeal radiographs in 100 Standardbred horses. Left and right proximal phalangeal foramina were symmetrical in 45 horses, but were asymmetrical in the remaining 55 horses. Of 200 proximal phalangeal foramina (in 100 horses), 78 were in the dorsal cortex, 61 were in the palmar cortex, and 61 were not visible radiographically. A significant ($P = 0.05$) effect of age or sex could not be determined. Dorsal nutrient foramina are those most commonly mistaken for fractures, presumably because of the length and vertical direction of the decreased density line. In 36 (63%) of 57 Standardbreds with dorsal nutrient foramina, these foramina were unilateral in the proximal forelimb phalanx. The authors concluded that radiographic comparison of the opposite limb would not have been a valid guideline for determining normality in more than half of these horses.

3. Reducing 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 erythrocyte 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 erythrocyte 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

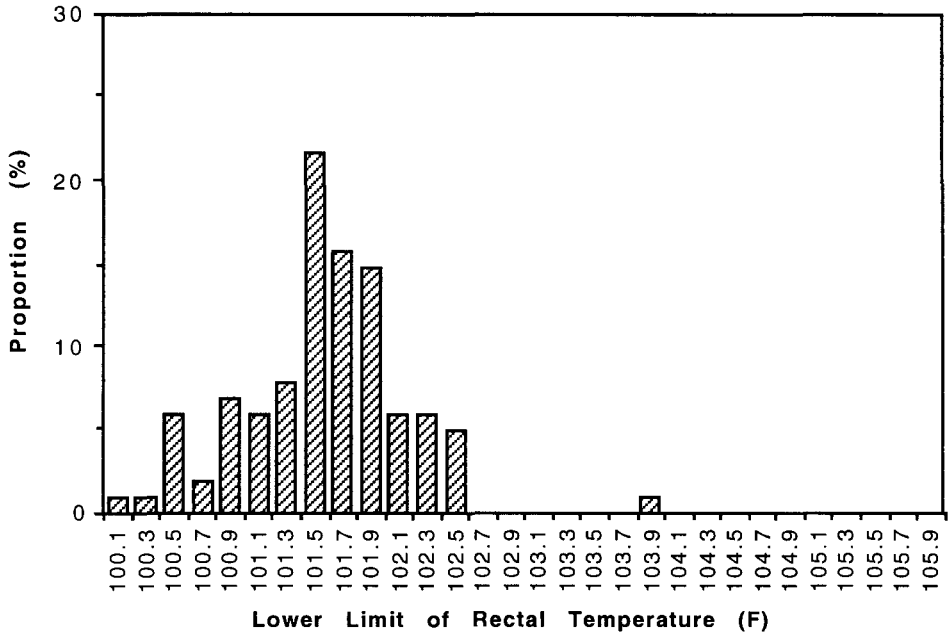


Figure 2.1 Frequency distribution of rectal temperatures in normal dogs.

and cats from birth to weaning. After that, they change drastically in the bovine and lymphocytes become the predominant peripheral leukocyte.

III. DISTRIBUTIONS

The adage that a picture is worth a thousand words (or numbers) is nowhere more true 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.

EXAMPLE: A typical histogram is depicted in Figure 2.1, which presents the distribution of 102 normal canine body temperatures. The size of each interval on the abscissa (x-axis) is 0.2^{o}F. We could have chosen any other interval, 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.

A. BASIC PROPERTIES OF DISTRIBUTIONS

Although Figure 2.1 represents a summary of 102 temperature readings, it is convenient to further summarize this 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 2.5 along with their advantages and disadvantages.

Table 2.5 Expressions of central tendency and dispersion for frequency distributions

Expression	*Definition*	*Advantages*	*Disadvantages*
		Measures of Central Tendency	
Mean	Sum of 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	Sometimes there are many "most frequent" values
		Measures of Dispersion	
Range	Lowest and highest values in a distribution	Includes all values	Greatly affected by extreme values
Standard deviation	The absolute value of the root mean squared deviation 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, quartile, quantile, etc.	The proportion of all observations falling between specified values	Describes the "usualness" of a value without assumptions about the shape of the distribution	Not well-suited for statistical manipulation

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

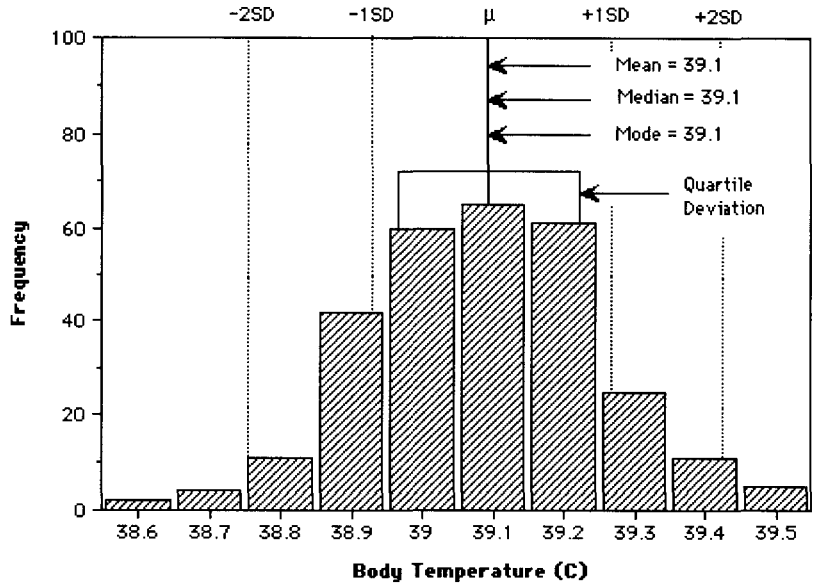


Figure 2.2 Frequency distribution of rectal temperature values for a cat over a 24-hour period. (Data courtesy of Dr. R.M. Weigel, College of Veterinary Medicine, University of Illinois. With permission.)

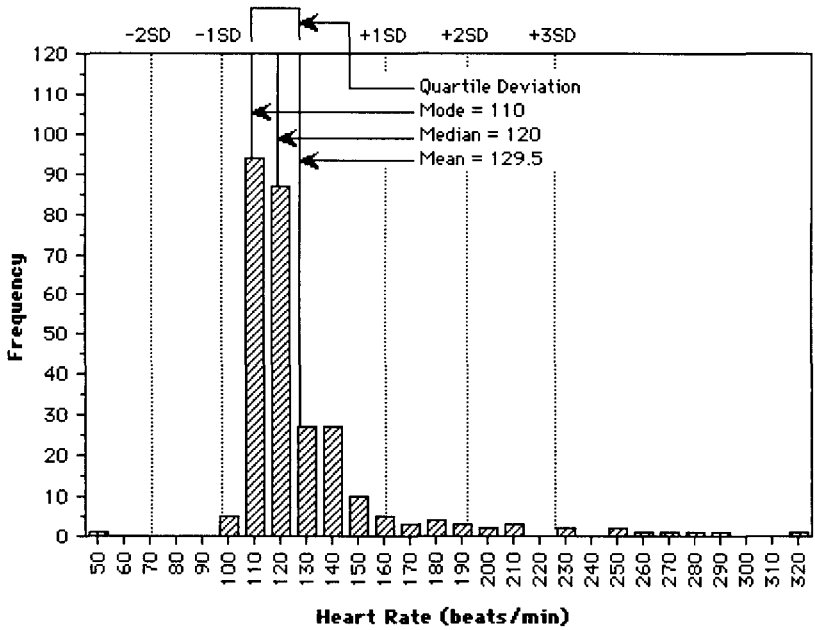


Figure 2.3 Frequency distribution of heart rate values for a cat over a 24-hour period. (Data courtesy of Dr. R.M. Weigel, College of Veterinary Medicine, University of Illinois. With permission.)

B. SHAPES OF NATURALLY OCCURRING DISTRIBUTIONS

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. Symmetry and Skewness

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 values of the distribution (often referred to as a "skewed to the right"). In negatively skewed distributions, the mean is less than the median, due to extreme values at the lower values of the distribution ("skewed to the left").

Figure 2.2 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^{o}C). Figure 2.3 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.

3. Factors Influencing the Shape of Frequency Distributions

Actual frequency distributions for many clinical measurements of animal populations change with characteristics such as age, sex, plane of nutrition and, in food-producing animals, stage of production.

EXAMPLE: Figure 2.4 depicts the frequency distribution of blood urea nitrogen (BUN) levels among 47 dairy herds (Payne et al, 1970). The data are only a portion of a battery of blood chemistry test results that were systematically collected from representative members of dairy herds to produce "metabolic profiles" (Stevens et al, 1980). The histograms actually represent the distribution of herd means. This is appropriate because the producer and veterinarian are often interested in herd performance rather than the health of individual animals. The metabolic profiles are used as a diagnostic aid to help identify metabolic problems in dairy herds that can then be corrected through improved feeding practices.

In this example the shape of the distribution curves vary with performance. For example, the BUN values of dry cows is broad and relatively flat, whereas that of middle-yield cows is skewed to the right. Thus, the timing and choice of population samples must be taken into consideration to avoid bias in test results.

C. 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 2.5) is a mathematical or theoretical model that describes the distribution of repeated measurements of the same physical properties 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 *approximate* a normal distribution, the latter is frequently used as a mathematical or theoretical model for calculating central

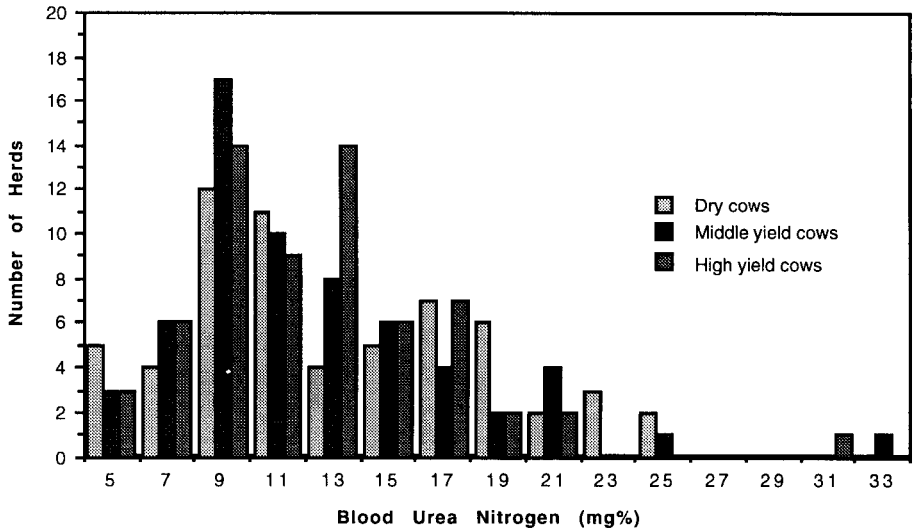


Figure 2.4 Distribution of results for urea in metabolic profile tests on 47 dairy herds. (From Payne, J.M., Dew, S.M., Manston, R. and Faulks, M. 1970. The use of a metabolic profile test in dairy herds. *Vet. Rec.* 87:150-157. With permission.)

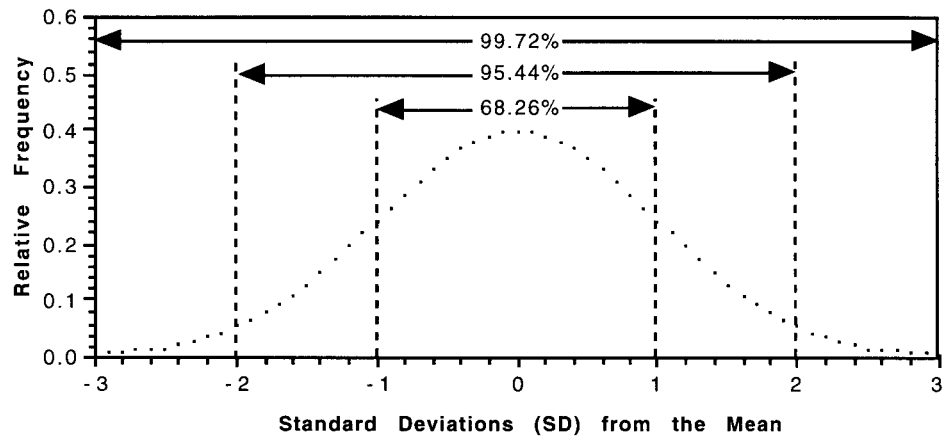


Figure 2.5 Percentages under the normal (Gaussian) curve at various standard deviations.

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

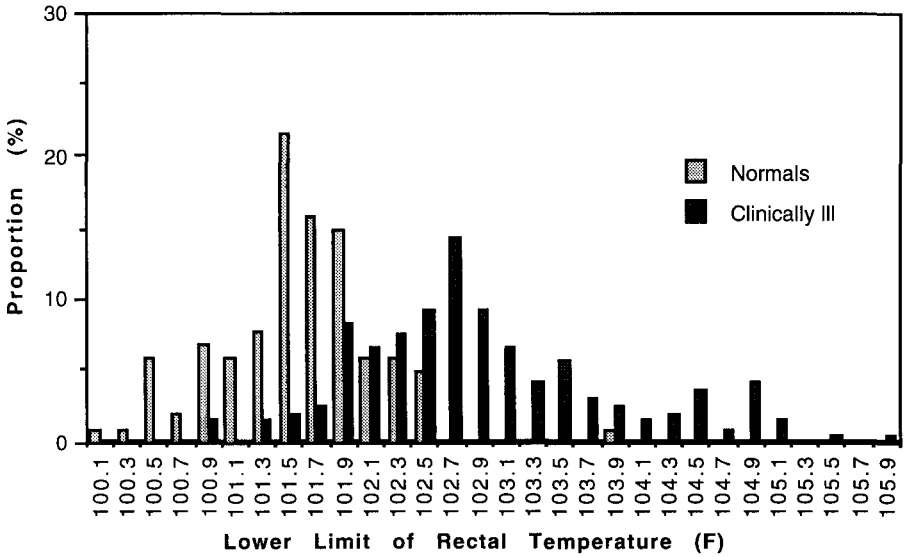


Figure 2.6 Frequency distribution of rectal temperatures for clinically normal and abnormal dogs.

approximately ±2 SDs from the mean, and approximately 99% of values will be within approximately ±3 SDs from the mean in a normal distribution.

IV. 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 abnormal? The distribution of clinical values among normal and diseased individuals frequently overlaps.

EXAMPLE: In Figure 2.6 the frequency distribution of body temperatures for a group of clinically normal dogs (from Figure 2.1) is superimposed on that for dogs exhibiting various signs of respiratory or gastrointestinal infection 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 proved useful: being unusual, being sick and being treatable."

When there is no clear division between normal and abnormal, three criteria have proved useful: being unusual, being sick and being treatable (Fletcher et al, 1982).

A. 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 cut-offs (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 2.1), the mean

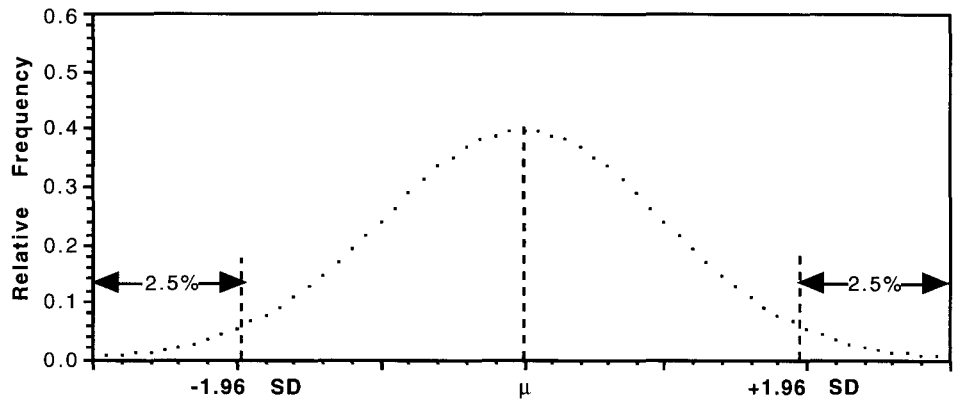


Figure 2.7 Mean (μ) and critical values (+/-1.96 SD) for the 95% confidence interval, under a ±two-tailed test of significance, where abnormality is associated with either high or low values, as blood leukocyte counts.

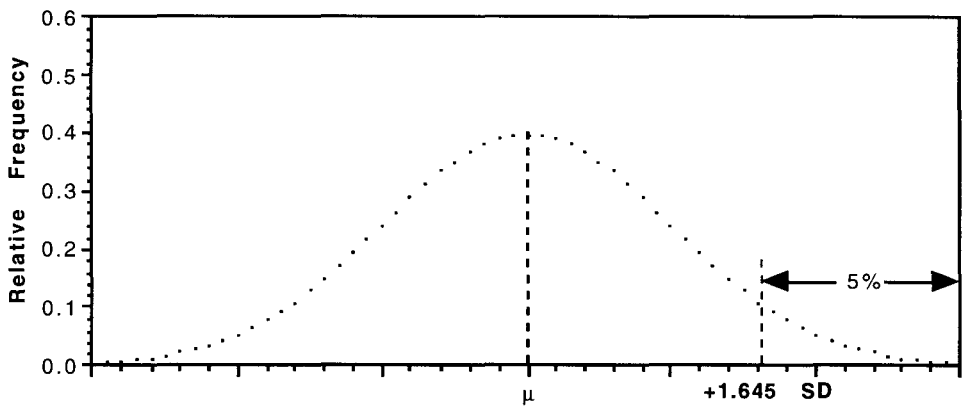


Figure 2.8 Mean (μ) and critical values (+1.645 SD) for the 95% confidence interval, under a one-tailed test of significance, where abnormality is associated with high values, as body temperatures.

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, corresponding to a *nondirectional hypothesis of normality* (Figure 2.7). 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 curve (Figure 2.8). 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. The one-tailed approach would yield a maximum normal temperature of approximately 102.6°F.

There are two limitations to the statistical approach to 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. These would be "false positive results" for the condition that we

Table 2.6 Percent of normal individuals expected to have at least one abnormal test result for a given number of tests using mean ±2 SD as the normal range

No. of Tests in Panel	*Percent with at Least One "Abnormal" Test*
1	4.5
2	8.9
3	13.1
4	17.0
5	20.8
6	24.4
7	27.9
8	31.2
9	34.3
10	37.3
11	40.2
12	42.9

From Boon, G.D. and Rebar, A.H. 1984. *Veterinary Values*, second edition. Schering Corp., Kenilworth, NJ. p. 7. With permission.

are measuring. If we were to extend the normal range to include 99% of the population, then the proportion of "false negative results" would increase, e.g., some truly abnormal individuals would be classified as normal. This concept becomes even more important 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 (Table 2.6).

The second important limitation is that mean and SD determinations assume that the results 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 2.1 approximate the normal distribution. Other data, such as canine plasma lactate values (Figure 2.9), do not. It is often assumed, as a matter of convenience, that clinical measurements are normally distributed.

EXAMPLE: A study was conducted to establish reference values for plasma lactate in beagle dogs (Evans, 1987). The frequency distribution of individual plasma lactate values for 60 healthy beagles is depicted in Figure 2.9. Note that the distribution is not normal and appears to be skewed to the right. The author did not define normal/abnormal cutoffs statistically for plasma lactate levels, but rather summarized the data in terms of mean (1.11 mmol/L) and range. The median, rather than the mean, would provide a better estimate of central tendency for these data, as half the population would be above and half below the median. The median would probably be lower than the mean.

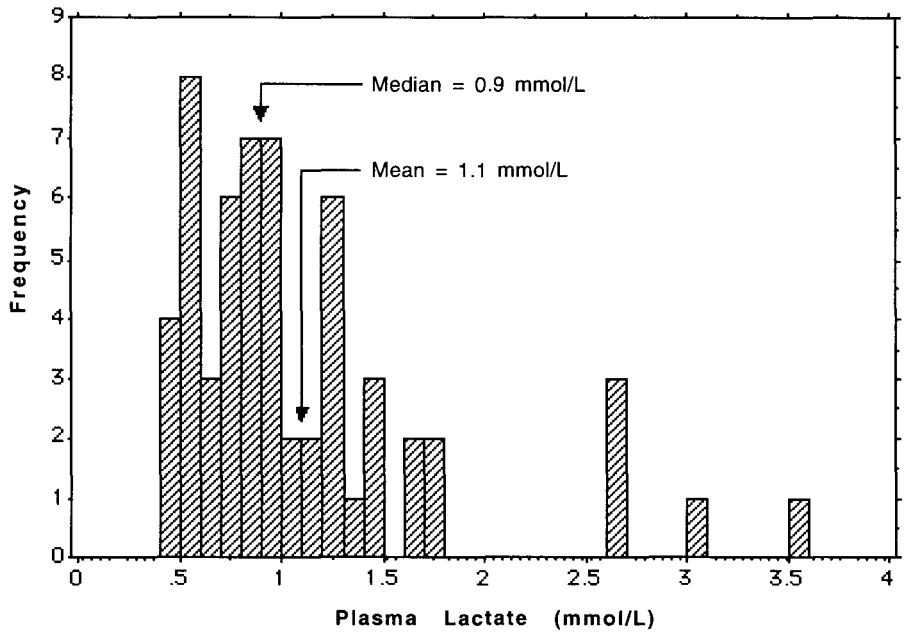


Figure 2.9 Frequency of plasma lactate values in 60 Beagle dogs. (From Evans, G.O. 1987. Plasma lactate measurements in healthy Beagle dogs. *Am. J. Vet. Res.* 48:131-132. With permission.)

Before making an assumption of normality, one should determine whether the distribution can in fact be approximated by a normal curve. This is usually done with a Chi-square goodness-of-fit test. If the normal values are not normally distributed, then one could express clinical values as a percentile of the cumulative distribution. This approach is independent of the shape of the distribution curve and provides an attractive alternative for determining the 95% cutoff. By this method the cutoff for the upper 2.5% of the normal distribution for canine temperatures (see Figure 2.1) would be 102.6°F.

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

B. 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 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 extensively in the next two chapters.

EXAMPLE: Table 2.7 presents the results of a study (House and Baker, 1968) designed to establish the normal/abnormal cutoff for serum gamma globulin levels based on risk of disease. Calves receive almost all of their maternal antibody by nursing rather than by transplacental transfer. Because serum gamma globulin levels are considered to indicate colostral absorption, studies by paper electrophoresis were made of the serum of 293 calves 3 to 6 days of age at calf-rearing units. The median percentage of gamma globulin for all calves was 12.1,

Table 2.7 Distribution of deaths and culls among calves according to percentage of serum gamma globulin

Group	*No. of Calves*	*Gamma Globulin (%)*	*Deaths*	*Culls*	*Total Loss*	*Loss (%)*
1	73	1.1-6.2	8	4	12	16.40
2	73	6.3-12.0	2	1	3	4.10
3	73	12.1-19.3	1	1	2	2.73
4	74	19.4-46.7	0	1	1	1.35
TOTAL	293		11	7	18	6.14

From House, J.A. and Baker, J.A. 1968. Comments on combination vaccines for bovine respiratory diseases. *J.A.V.M.A.* 152:893-894. With permission.

with a range of 1.1% to 46.7%. The percentage of gamma globulin in experimentally deprived calves is reported to range from 1.5% to 3.0%. As normal values had not been established, the calves were allotted to four equal-sized groups (quartiles) based on the percentage of gamma globulin, and their performance was monitored. The results show that the percentage of loss (deaths and culls) increases as the percentage of gamma globulin decreases, and that gamma globulin levels below approximately 7% should be considered abnormal. Application of these criteria would result in 25% of calves being considered abnormal versus only 5% using the statistical approach described previously.

C. ABNORMAL AS TREATABLE OR DETECTABLE

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, culminating in the recently introduced avermectins, 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 medicine abnormality may be defined as the point at which treatment is economically justified. This point, termed the *economic threshold*, is dependent on the cost of treatment and the economic gain that can be expected. To be effective in these situations, a veterinarian must be knowledgeable in economic analysis as well as in medicine.

V. SUMMARY

The process of medical decision-making consists of four components: collection of (1) subjective and (2) objective data, (3) assessment of the situation, and finally (4) a plan of action. There are three principal scales used for measuring clinical phenomena: nominal, ordinal and interval. Nominal data can be placed into discrete categories that have no inherent order. Another name for nominal data is categorical data. Ordinal data can be ranked, but the intervals are not uniform in size. Data that are ordered and for which the size of the intervals are known are called interval.

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. Validity and reliability are more difficult to establish for other clinical measurements that rely on our senses and for which no physical standards exist.

There are two major sources of variation in clinical measurements. Measurement variation is associated with the act of measurement itself and may be due to the performance of the instruments being used, the observers themselves, or both. Biological variation can manifest at all levels of an animal population. As a rule, rigid adherence to test protocols is the single most important way to reduce overall test variation.

Two basic properties of distributions can be used to summarize interval data: central tendency, or the middle of the distribution, and dispersion, an index of the spread of the data. The most common measures of central tendency and dispersion are the mean and SD, respectively. The frequency distribution for a variable can have one or more measurement values with the maximum frequency, or mode. 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. 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, while in negatively skewed distributions, the mean is less than the median due to extreme values at the lower range of the distribution ("skewed to the left").

Actual frequency distributions for many clinical measurements of animal populations change with characteristics such as age, sex, plane of nutrition and, in food-producing animals, stage of production. The normal distribution is a mathematical or theoretical model which represents random variation alone. It is frequently used to estimate the limits of normality.

Three criteria that have been used to distinguish normal from abnormal are (1) being unusual, (2) being sick and (3) being treatable. Being unusual assumes that normal values are distributed normally and that values outside of the normal range, defined as the mean +/-2 SD (for a two-tailed test of significance), are abnormal. One disadvantage of this approach is that approximately 5% of normal individuals would be classified as abnormal on any single test. Another disadvantage is that natural distributions may not conform to the normal distribution. Being sick relies on calling abnormal those findings that are regularly associated with disease, disability, unproductivity, or death. Being treatable defines abnormal as the level that is worth treating.

Chapter 3

EVALUATION OF DIAGNOSTIC TESTS

I. 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 frequency 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 on 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 tests. *Diagnostic tests* are used to distinguish between animals that have the disease in question and those that have other diseases on the differential list (White, 1986). Diagnostic testing begins with diseased individuals. *Screening tests* are used for the presumptive identification of unrecognized disease or defect in apparently healthy populations. Screening tests begin 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 for the evaluation of diagnostic tests are summarized in Table 3.1.

A distinction must be made between diagnostic and screening tests. Diagnostic testing begins with diseased individuals, whereas screening tests begin with presumably healthy individuals.

II. TEST ACCURACY

Test accuracy is the proportion of all tests, 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.

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 (Dubensky

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: Bayes' graph	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*	Testing and treatment thresholds

*See Chapter 14.

and White, 1983). As a result, diagnostic testing is frequently approached in stages, substituting simpler tests for more rigorous ones, at least initially.

EXAMPLE: The diagnostic strategy for tumors of the mammalian lymphoid and hemopoietic tissues includes several tests varying in cost and accuracy. These tumors include canine malignant lymphoma, feline lymphosarcoma and leukemia, and bovine leukosis. For example, bovine leukosis may initially be suspected based on relatively nonspecific evidence such as unthriftiness, visual swelling of lymph nodes, morphologic appearance of circulating leukocytes, and changes in blood biochemical parameters. A serologic test for bovine leukosis virus (BLV) infection may next be performed to ensure that the animal in question has been exposed to the virus, thus increasing the likelihood that the animal is truly suffering from BLV. Finally, a lymph node biopsy may be performed to determine the true cause of lymph node enlargement. The proof of the diagnosis, or "gold standard," will come after you have convinced the owner of your diagnosis and a necropsy is performed.

For economic reasons, the diagnostic strategy for the avian leukosis complex, or Marek's disease, a similar neoplastic disease of poultry, would be quite different. Because the economic value of individual birds is insignificant, a sample of afflicted birds from the flock would be necropsied immediately to determine the disease status of the flock.

A. 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 refers to the means by which one can determine whether a disease is truly present or not. 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 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 (Holden, 1985). However, many disorders cannot be confirmed even at necropsy, because they stem from subtle biochemical or neurologic alterations measurable only in the living animal.

B. POSTMORTEM EXAMINATION AS A DIAGNOSTIC TEST

Postmortem examination is used more frequently as a diagnostic tool in veterinary medicine than in human medicine. In fact, the proportion of human deaths in the United States followed by autopsy has declined from nearly 50% shortly after World War II to the current level of approximately 15% of all deaths (Geller, 1983; Holden, 1985). Even with the many recent advances in human medicine, studies have revealed that the major diagnosis has been wrong in as many as 40% and the immediate cause of death clinically unrecognized in more than 40% of people autopsied. This error rate, when coupled with the small percentage of autopsies performed, implies that at least half of the 2 million death certificates recorded in the United States each year are in error. Consequently, it is believed that the value of human health statistics dealing with the causes of death has deteriorated (Geller, 1983).

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.

Slaughter checks are already part of the diagnostic and surveillance programs performed by food animal practitioners for their clients (Anonymous, 1985a; Straw, 1985; Straw et al, 1986). The use of data generated during routine activities of the Food Safety Inspection Service (FSIS) represents an attempt to include carcass inspection data in a disease surveillance program (Houston, 1984; King, 1985). Components of the program include (1) risk-based allocation of inspection resources, (2) statistically based sampling strategies and (3) a livestock and poultry disease reporting system.

III. 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.

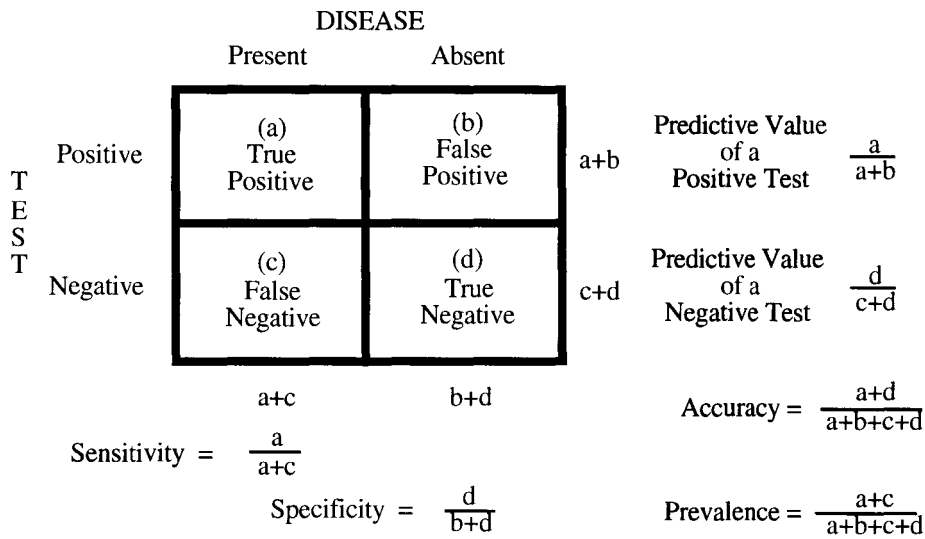


Figure 3.1 Diagnostic test outcomes and definitions. There are four possible test outcomes: two are correct and two are incorrect. Values for all four outcomes are used to estimate test sensitivity, specificity, predictive value and accuracy and the prevalence of disease in the population.

A. SENSITIVITY AND SPECIFICITY (TRUE-POSITIVE AND -NEGATIVE RATES)

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

EXAMPLE: Case series are excellent sources of data on the sensitivity of a particular test or finding. The frequency of clinicopathologic findings associated with chronic renal disease in cats in Table 3.2 (DiBartola et al, 1987) demonstrates the effect of biological variation on test sensitivity. Sensitivity data such as these provide useful criteria for ruling out diseases on a differential list. For example, among serum biochemical findings, azotemia was present in 97% of affected cats, whereas hyperchloremia was present in only 3.2%. Thus, if a patient were presented with clinical signs suggestive of chronic renal disease (lethargy, anorexia, weight loss), normal blood creatinine levels would provide a better basis for ruling out the diagnosis than would normal chloride levels. One caveat in this study is that it is not clear how chronic renal disease was confirmed in the cats (gold standard).

B. 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 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).

Table 3.2 Hematologic and serum biochemical findings in cats with chronic renal disease

Clinicopathologic Finding	*% of Cats*	*Clinicopathologic Finding*	*% of Cats*
Hematologic findings		*Biochemical findings (continued)*	
Hyperproteinemia (>8.0 g/dl)	61.6	Hypokalemia (<3.6 mEq/L)	29.7
Lymphopenia (<1200/μl)	56.9	Hyponatremia (<149 mEq/L)	29.7
Nonregenerative anemia (PCV<27%)	41.1	Hyperglycemia (>125 mg/dl)	23.5
Leukocytosis (>20,000/μl)	27.4	Increased anion gap (>35 mEq/L)	18.6
Leukopenia (<6000/μl)	4.1	Hypocalcemia (<8.3 mg/dl)	14.8
Hypoproteinemia (<6.0 g/dl)	2.7	Hypercalcemia (>10.5 mg/dl)	11.5
Biochemical findings		Hypoalbuminemia (<2.3 g/dl)	11.1
Azotemia (creatinine>1.8 mg/dl)	96.9	Hyperalbuminemia (>3.6 g/dl)	9.3
(BUN>35 mg/dl)	95.8	Hypernatremia (>162 mEq/L)	7.8
Hypercholesterolemia (>155 mg/dl)	72.5	Hyperkalemia (>5.4 mEq/L)	6.2
Decreased CO_2 combining power (<15 mEq/L)	62.7	Hypochloremia (<105 mEq/L)	4.8
Hyperphosphatemia (>7.1 mg/dl)	58.3	Hyperchloremia (>135 mEq/L)	3.2

From DiBartola, S.P., Rutgers, H.C., Zack, P.M., and Tarr, M.J. 1987. Clinicopathologic findings associated with chronic renal disease in cats: 74 cases (1973-1984). *J.A.V.M.A.* 190:1196-1202. With permission.

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.

EXAMPLE: The significance of the comparisons in Figure 3.1 can be appreciated by inserting data for an enzyme-linked immunosorbent assay (ELISA) test for antibody to *Mycobacterium paratuberculosis*, causative agent of paratuberculosis, or Johne's disease of cattle (Figure 3.2). The true infection status of the cattle (gold standard) was determined by fecal culture (Spangler et al, 1992). Serologic test sensitivity was 72.9%. The 27.1% of infected cattle that were not detected are referred to as false-negatives. Serologic test specificity was 84.8%, with 15.2% false-positive results.

Figure 3.3 depicts the same data as a frequency polygon in which the frequency of ELISA values for fecal culture-negative and -positive cattle is related to the cutoff value of 0.35. Any

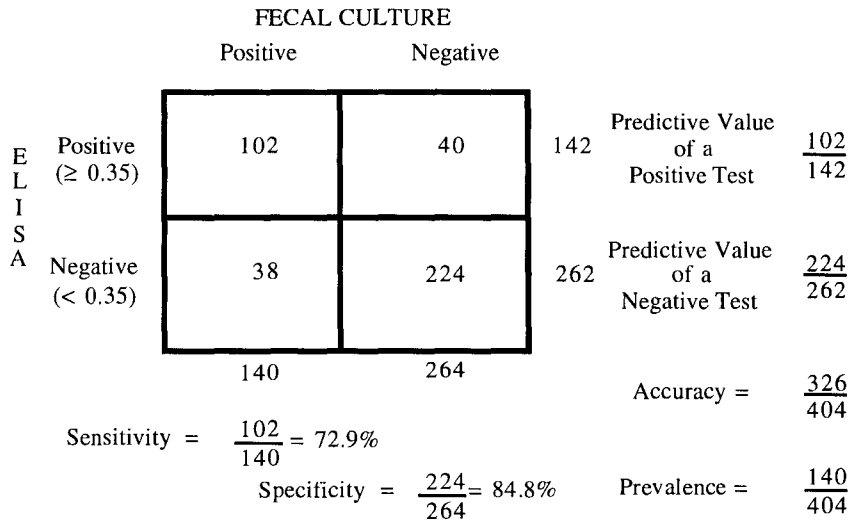


Figure 3.2 Evaluation of an enzyme-linked immunosorbent assay (ELISA) for the detection of antibody to *Mycobacterium paratuberculosis*. In this example, any ELISA value ≥ 0.35 (≥35% of the optical density of the positive reference serum) is considered positive, and any value < 0.35 is considered negative. (Source of data: Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cut-offs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204. With permission.)

shift in the ELISA cutoff criterion to the left or right would necessitate a recalculation of test parameters summarized in Figure 3.2.

C. 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 are absolute properties of a test and do not change for any given cutoff value, predictive values are relative, varying with the prevalence of disease in the population from which the patient came. For a full discussion of prevalence, see Chapter 5.

D. THE EFFECT OF PREVALENCE ON PREDICTIVE VALUES

Diagnostic tests are used in populations with widely differing disease frequencies. As indicated previously, this 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.

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. Therefore, improved sensitivity and specificity cannot be expected to result in a dramatic improvement in predictive value.

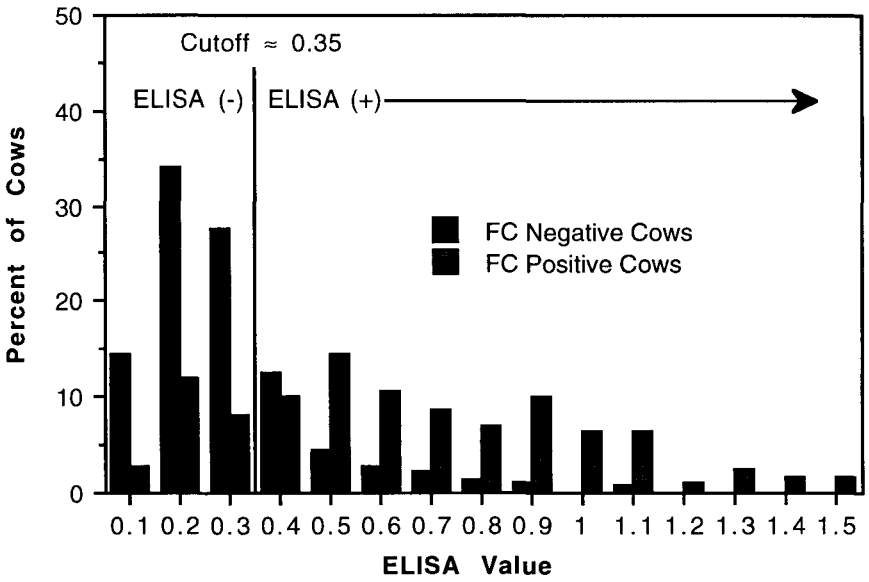


Figure 3.3 Frequency distribution of ELISA values for fecal culture (FC)-negative and -positive cattle summarized in Figure 3.2. Any ELISA value ≥ 0.35 (≥35% of the optical density of the positive reference serum) is considered positive, and any value < 0.35 is considered negative. (Source of data: Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cutoffs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204. With permission.)

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, such as the bovine brucellosis eradication program. Use of a serologic test of low specificity (and therefore low positive predictive value) could, in theory, lead to depopulation of the entire herd.

EXAMPLE: 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 intensity 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.

E. LIKELIHOOD RATIOS

The likelihood ratio is an index of diagnostic utility that expresses the odds 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 (Sackett, 1992). By "finding" we

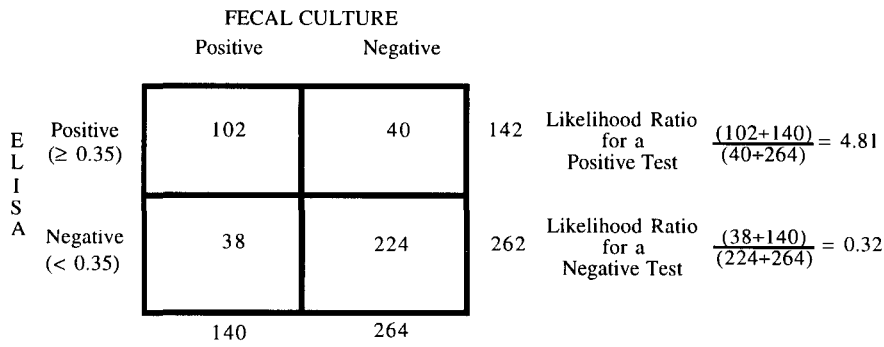


Figure 3.4 Calculation of positive and negative likelihood ratios from data presented in Figure 3.2 on an ELISA test for *M. paratuberculosis* antibody in cattle. The likelihood ratio for a positive test (≥ cutoff) = sensitivity ÷ (1 - specificity), or true-positive rate ÷ false-positive rate. The likelihood ratio for a negative test (< cutoff) = (1 - sensitivity) ÷ specificity, or false-negative rate ÷ true-negative rate. (Source of data: Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cut-offs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204. With permission.)

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 using the same four values used to calculate other aspects of test performance (Figure 3.4).

The likelihood ratio for a positive test is the ratio of the true-positive rate (pT+/D+) divided by the false-positive rate (pT+/D-), or equivalently, sensitivity/(1 - specificity). The likelihood ratio for a negative test is the ratio of the false-negative rate (pT-/D+) divided by the true-negative rate (pT-/D-), or equivalently, (1 - sensitivity)/specificity. The ideal diagnostic test would yield a likelihood ratio of infinity for a positive test (e.g., 100%/0%) and a likelihood ratio of 0 for a negative test (e.g., 0%/100%). A likelihood ratio of one for either a positive or negative test means the test result conveys no information. In the paratuberculosis test example shown in Figure 3.4, the likelihood ratio for a positive test is 4.81 (72.86%/15.15%), meaning that an ELISA value ≥ 0.35 is almost five times as likely to have come from an *M. paratuberculosis-infected* versus *-uninfected* animal. The likelihood ratio for a negative test is 0.32 (27.14%/84.85%), meaning that an ELISA value < 0.35 is about one-third as likely to have come from an *infected* versus *uninfected* animal.

The likelihood ratio offers several advantages over other methods of reporting test performance. Because the likelihood ratio is derived from test sensitivity and specificity only, it is unaffected by disease prevalence, making it an especially stable expression of test performance. 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. For example, by expanding the levels of *M. paratuberculosis* test results from two (as in the 2 x 2 table above) to ten (as in Table 3.4) the range of likelihood ratios has widened from 15-fold (0.32 to 4.81) to 327-fold (0.15 to 49.03). 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 expressed in terms of a single positive/negative cutoff. Finally, the likelihood ratio can be used to estimate the actual probability of any disease on a differential list, if its pretest probability is known. This application of the likelihood ratio will be discussed in the next chapter.

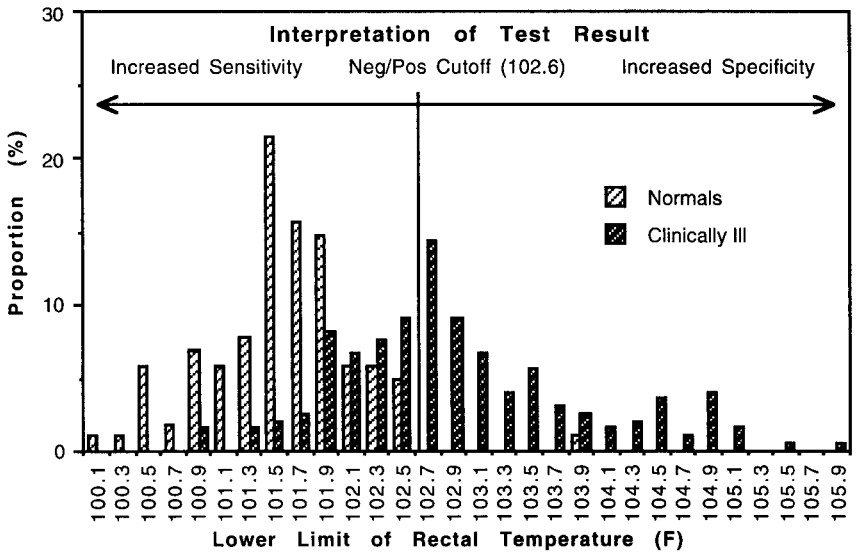


Figure 3.5 Frequency distribution of rectal temperatures from normal and abnormal dogs to demonstrate the effect of moving the negative/positive cutoff on the sensitivity and specificity of a diagnostic test.

F. ACCURACY, REPRODUCIBILITY AND CONCORDANCE

Accuracy, reproducibility and concordance are other terms used to describe diagnostic test performance. *Accuracy* is estimated directly from the same 2 by 2 table used to estimate other test properties and is the proportion of all tests, both positive and negative, that are correct (see Figure 3.1). It 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 accuracy of the test (except for the special condition where test sensitivity and specificity are equal).

Reproducibility refers to the degree to which repeated tests on the same sample(s) give the same result (see Validity and Reliability, Chapter 2), 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.

EXAMPLE: Schwartz et al (1989) evaluated the interlaboratory and intralaboratory agreement of Lyme disease test results among four independent laboratories for serum specimens from 132 outdoor workers in New Jersey. The measurement of agreement employed, the kappa statistic, ranged from 0.45 to 0.53 among the four laboratories, representing low levels of agreement. Of 20 sera reported as positive by at least one laboratory, 85%, 50% and 30% were reported positive by two, three and four laboratories, respectively. The kappa statistic is discussed in Chapter 9.

IV. INTERPRETATION OF TESTS WHOSE RESULTS FALL ON A CONTINUUM

A. 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 3.5 depicts the distribution of rectal temperatures for the two popula-

Table 3.3 Effect of cutoff on the performance of an ELISA test for *Mycobacterium paratuberculosis* infection in cattle

	Fecal Culture						
*ELISA Cutoff**	*Number Positive*	*Number Negative*	*Sensitivity (%)¶*	*Specificity (%)†*	*False Neg§*	*False Pos¥*	*Sum*
<10	3	39	100	0	0	264	264
10	16	91	98	15	3	225	228
20	11	73	86	49	19	134	153
30	14	33	79	77	30	61	91
40	20	11	69	89	44	28	72
50	15	7	54	94	64	17	81
60	12	5	44	96	79	10	89
70	9	3	35	98	91	5	96
80	14	1	29	99	100	2	102
≥90	26	1	19	99.6	114	1	115
Totals	140	264					

*ELISA values expressed as a percent of the optical density of the positive reference serum.

$$\text{¶Sensitivity} = \frac{\text{No. ELISA(+)/fecal culture(+)} \geq \text{cutoff}}{\text{total fecal culture(+)}}$$

$$\text{†Specificity} = \frac{\text{No. ELISA(+)/fecal culture(-)} < \text{cutoff}}{\text{total fecal culture(-)}}$$

§Number of false negative diagnoses at each cutoff = (140) x (1 - sensitivity).
¥Number of false positive diagnoses at each cutoff = (264) x (1 - specificity).

Source of data: Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cutoffs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204. With permission.

tions of dogs discussed earlier (see Figure 2.6), 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.

EXAMPLE: Spangler et al (1992) evaluated the ability of different cutoffs in a quantitative ELISA to discriminate between *Mycobacterium bovis*-infected and -uninfected cattle. One hundred and forty cows with fecal culture-confirmed infection served as cases, while 264 fecal culture-negative cattle were controls. The sensitivity of the ELISA in diagnosis of *M. bovis* infection decreased from 100% to 19% as the cutoff value for a positive test (as a percent of

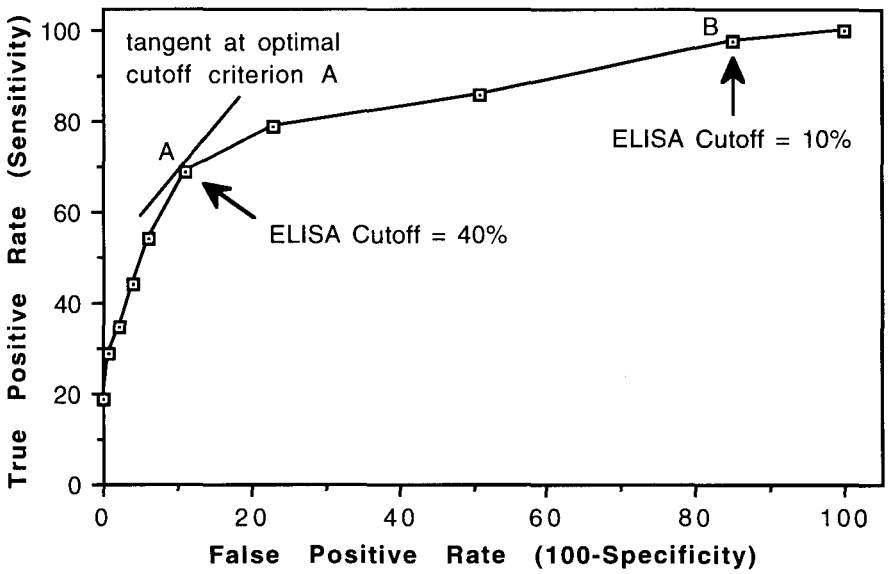


Figure 3.6 Response-operating characteristic (ROC) curve for an enzyme-linked immunosorbent assay (ELISA) for the diagnosis of *Mycobacterium paratuberculosis* infection in cattle. Points A and B identify optimum cutoff points when (A) the cost of a false negative = the cost of a false positive, and (B) the cost of a false negative is ten times that of a false positive. Corresponding ELISA values are approximately 40% and 10%. See Table 3.3 for corresponding sensitivity and specificity values. (Source of data: Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cutoffs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204. With permission.)

the OD of the positive reference serum) was increased from <10% to ≥90%, while specificity increased from 0% to 99.6% over the same range of cutoff values (Table 3.3). Regardless of the cutoff point, a risk of misdiagnosis will always exist. In this example, increasing the cutoff decreases test sensitivity but increases specificity. Decreasing the cutoff has the opposite effect.

The effect of test sensitivity and specificity on predictive values can be appreciated by studying the number of false-negative and false-positive diagnoses in Table 3.3, which illustrate an important point: tests of low sensitivity increase the likelihood of false-negative results, whereas tests of low specificity increase the likelihood of false-positive test results.

B. RESPONSE-OPERATING CHARACTERISTIC (ROC) CURVES

For test results that fall along a continuum, e.g., ELISA cutoffs for *M. paratuberculosis* infection (Table 3.3), test performance can be depicted graphically by plotting a response-operating characteristic (ROC) curve (also variably called a receiver- or relative-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. The ROC curve provides a simple method for evaluating a test's ability to discriminate between health and disease over the complete spectrum of operating conditions, and it can be used to select cutoffs (decision thresholds) or to compare diagnostic tests.

An ROC curve for the data in Table 3.3 is depicted in Figure 3.6. Each point on the ROC curve defines a set of operating characteristics for the test based on sensitivity and specificity. The astute reader will note that the ROC curve is really only a series of likelihood ratios, 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 assessing and using diagnostic tests (Zweig and Campbell, 1993).

C. 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 should include a consideration of (1) the distribution of results in two different populations - normal patients and patients with disease, (2) the prevalence of disease in the population to be tested, and (3) the costs 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 the ROC curve closest to the upper left-hand corner, where test sensitivity and specificity are maximized, e.g., (sensitivity + specificity)/2 attains its highest value (Sackett et al, 1991). In the *M. paratuberculosis* example (prevalence = 34.7%), the lowest total number of diagnostic errors occurs at an ELISA cutoff of approximately 40% (Table 3.3).

This approach does not take into account the relative cost of false-positive versus false-negative diagnoses, which can be factored in by simply multiplying the relative or absolute cost by the respective number of false-negative and false-positive diagnoses and summing the result for each cutoff. Alternatively, the optimum cutoff can be determined from the ROC curve by identifying the point where the slope of the ROC curve equals

$$\frac{pD- \times CFP}{pD+ \times CFN}$$

where CFP is the cost of a false-positive, CFN is the cost of a false-negative, and pD+ and pD- are the proportion of diseased and healthy animals, respectively, in the test population (McNeil et al, 1975). Application of this formula selects the lowest total cost for false-positive and false-negative tests combined.

In the *M. paratuberculosis* example (Figure 3.4), if false-negative and false-positive errors were considered equally undesirable, then the optimal operating point on the ROC curve at the observed prevalence of 34.7% would have a slope of [(0.653 x 1) ÷ (0.347 x 1)] or 1.882, corresponding to a normal/abnormal ELISA cutoff of approximately 40% (criterion A in Figure 3.6). Test sensitivity would be 69% and specificity would be 89% (Table 3.3). If, on the other hand, a false-negative error were considered to be five times as bad as a false-positive error (e.g., a significant penalty for failing to detect the disease), then the optimal cutoff point would have a slope of [(0.653 x 1) ÷ (0.347 x 10)] or 0.188, corresponding to an ELISA cutoff of approximately 10% (criterion B in Figure 3.6). Test sensitivity would be 98% and specificity would be 15%. In this example we are willing to accept the relatively high rate of

Table 3.4 Relationship between ELISA optical density (OD) and likelihood of fecal shedding of *Mycobacterium bovis* in cattle.

	Fecal Culture		*Likelihood Ratio*	
ELISA Cutoff	*Number Positive*	*Number Negative*	*Between Cutoffs**	*≥ Cutoff†*
<10	3	39	0.15	1.00
10	16	91	0.33	1.15
20	11	73	0.28	1.70
30	14	33	0.80	3.40
40	20	11	3.43	6.47
50	15	7	4.04	8.43
60	12	5	4.53	11.50
70	9	3	5.66	18.48
80	14	1	26.40	37.71
≥90	26	1	49.03	49.03
Totals	140	264		

ELISA values expressed as a percent of the optical density of the positive reference serum.

*Likelihood ratio between cutoffs =

$$\frac{\text{No. ELISA(+)/fecal culture(+) between cutoffs} \div \text{total fecal culture(+)}}{\text{No. ELISA(+)/fecal culture(-) between cutoffs} \div \text{total fecal culture(-)}}$$

†Likelihood ≥ cutoff =

$$\frac{\text{No. ELISA(+)/fecal culture(+)} \geq \text{cutoff} \div \text{total fecal culture(+)}}{\text{No. ELISA(+)/fecal culture(-)} \geq \text{cutoff} \div \text{total fecal culture(-)}}$$

Source of data: Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cutoffs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204. With permission.

false-positive results because of the ten-fold greater penalty for a false-negative result. If the cost differential or prevalence of TRP were to change, then so would the optimal cutoff.

D. LIKELIHOOD RATIO ANALYSIS

Another question that one may ask about test results that fall on a continuum is whether there is a correlation between the magnitude of the test result and the likelihood (probability) of disease. The likelihood ratio can be used to test the strength of this relationship. Returning to the data of Spangler et al (1992), likelihood ratios have been calculated over the range of ELISA values registered by animals who were shedding or not shedding *M. paratuberculosis* in their feces (Table 3.4). Likelihood ratios were calculated in two ways: (1) based on the number of shedders and nonshedders whose ELISA values fell between ELISA cutoff values and (2) based on the number of shedders and nonshedders whose ELISA values were ≥ a given cutoff value.

Likelihood ratio analysis revealed that there was a strong positive correlation between the amount of serum antibody to *M. paratuberculosis* (based on ELISA optical density) and the

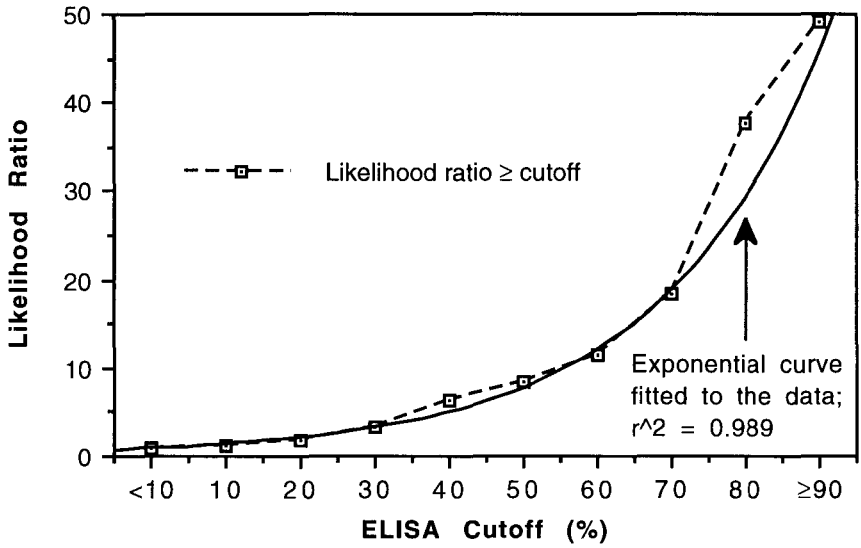


Figure 3.7 Likelihood ratio analysis of data from Table 3.4. Likelihood ratios were based on the proportion of *M. paratuberculosis* fecal culture-positive and -negative cows whose ELISA values were ≥ a given cutoff. The solid line is an exponential curve fitted to the data. The high r^2 value (0.989) suggests a strong positive correlation between the amount of serum antibody to *M. paratuberculosis* and the likelihood of fecal shedding. (Source of data: Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cut-offs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204. With permission.)

likelihood of fecal shedding (Figure 3.7). An exponential equation provided the best fit to the data ($r^2 = 0.989$), which is probably a reflection of the way in which ELISA values were expressed, e.g., as a proportion of the positive control. Knowing the magnitude of the ELISA OD value would be useful not only for interpreting a test result as positive or negative, but also for prioritizing cows for culling from the herd. The use of the likelihood ratio for estimating the actual probability of disease is discussed in the next chapter.

V. COMPARISON OF DIAGNOSTIC TESTS

A. FOR 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. The mathematical relationship among pretest and post-test probabilities and test results was described hundreds of years ago in *Bayes' theorem*. 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. Test accuracies can be compared using a Bayesian graph in which the post-test probability of disease, given a positive or negative test result, is plotted over a range of pretest probabilities (prevalences) from 0 to 100%. This approach effectively simulates a test's performance over its entire range of operating conditions, whether used as a screening test (low pretest probability of disease) or as a diagnostic test

(high pretest probability of disease). Bayes' formulas for estimating the post-test probability of disease given a positive or negative test result are useful for this analysis.

The post-test probability of disease given a *positive* test equals

$$\frac{\text{true-positives}}{\text{all positives}} = \frac{\text{pD x sensitivity}}{\text{pD x sensitivity} + [(1\text{-pD}) \text{ x } (1 - \text{specificity})]}$$

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

$$\frac{\text{false-negatives}}{\text{all negatives}} = \frac{\text{pD x } (1 - \text{sensitivity})}{\text{pD x } (1 - \text{sensitivity}) + [(1\text{-pD}) \text{ x (specificity)}]}$$

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

An example of the application of Bayesian formulas is described in the following example where three tests for the detection of intramammary infection of cows with *Staphylococcus aureus* are compared. The lines in Figure 3.8 correspond to the likelihood of disease given a positive or negative test result for each of the tests. The best tests are those that generate lines of greatest curvature. Tests whose post-test probabilities fall close to the straight diagonal line convey no useful additional information.

EXAMPLE: Management of bovine mastitis caused by *Staphylococcus aureus* focuses primarily on preventing the spread of the bacterium within the herd rather than treating individual animals. Infections with *S. aureus* are often subclinical, making identification of cows with intramammary infections difficult. The most sensitive testing strategy for detecting intramammary *S. aureus* infection is microbiologic culture of several consecutive milk samples. However, this approach is costly and time-consuming, prompting the search for simpler diagnostic tests.

Hicks et al (1994) compared microbiologic culture, somatic cell counts (SCC), and an enzyme-linked immunosorbent assay (ELISA) performed on single milk samples as methods of identifying cows with *S. aureus* intramammary infection. Cows were considered positive for *S. aureus* (gold standard) if *S. aureus* was isolated from at least two of the first three consecutive milk samples collected from that cow. Test results were interpreted as follows:

- Microbiologic culture - Any cow with ≥1 coagulase-positive colony in a milk sample from any quarter was classified as culture-positive for *S. aureus*.
- Somatic cell count - Cows were classified as high SCC (positive) if there were ≥200,000 cells/ml of milk sample and low SCC (negative) if there were <200,000 cells/ ml.
- ELISA test - Cows were classified as positive if the optical density (OD) in the test well was greater than that of the positive control well and negative if the OD was less than or equal to that of the positive control.

The tests are compared in Figure 3.8 where the post-test probability of *S. aureus* infection given a positive or negative test is plotted over a range of pretest probabilities from 0 to 100%. The differences among the tests with regard to test sensitivity are reflected in progressively greater deviation from linearity of the respective curves for negative test results. Because of the greater sensitivity of microbiologic culture, we can be more confident that a cow with a negative test result is free of infection than with a negative result using the other two tests. The differences in test specificity are reflected in the respective curves for positive

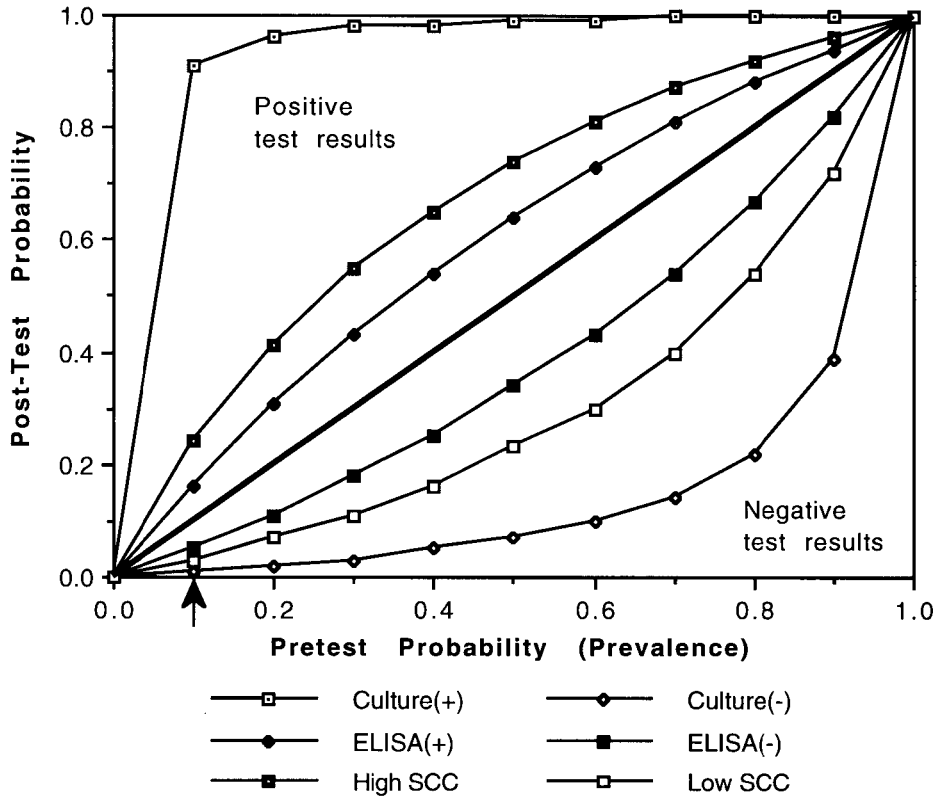


Figure 3.8 Comparison of the pretest and post-test probabilities of *Staphylococcus aureus* intramammary infection in dairy cows for three single-sample milk screening tests. Curves were drawn using Bayes' formula for the post-test probability of disease given a positive or negative test result (see above and next chapter). Test sensitivities and specificities are, respectively, microbiologic culture (93%, 99%), ELISA (69%, 61%) and somatic cell count (79%, 72%). (Source of data: Hicks, C.R., Eberhart, R.J., and Sischo, W.M. 1994. Comparison of microbiologic culture, an enzyme-linked immunosorbent assay, and determination of somatic cell count for diagnosing *Staphylococcus aureus* mastitis in dairy cows. *J.A.V.M.A.* 204:255-260. With permission.)

test results. Because of the greater specificity of microbiologic culture, a positive result provides stronger evidence of *S. aureus* infection than does a positive result with the other two tests.

Highly sensitive tests are most useful when their results are negative, and highly specific tests are most useful when their results are positive.

Because these tests are intended to be used as screening tests, comparing test performance at low prevalence rates is especially important. At a prevalence rate of 10% (arrow), the accuracy of a negative test result is comparable for all three tests. However, only microbiologic culture provides any useful information from a positive test result at low prevalence rates. It is clear that microbiologic culture of a single milk sample is far more accurate than either somatic cell counts or ELISA for identifying cows with *S. aureus* intramammary infections. Figure 3.8

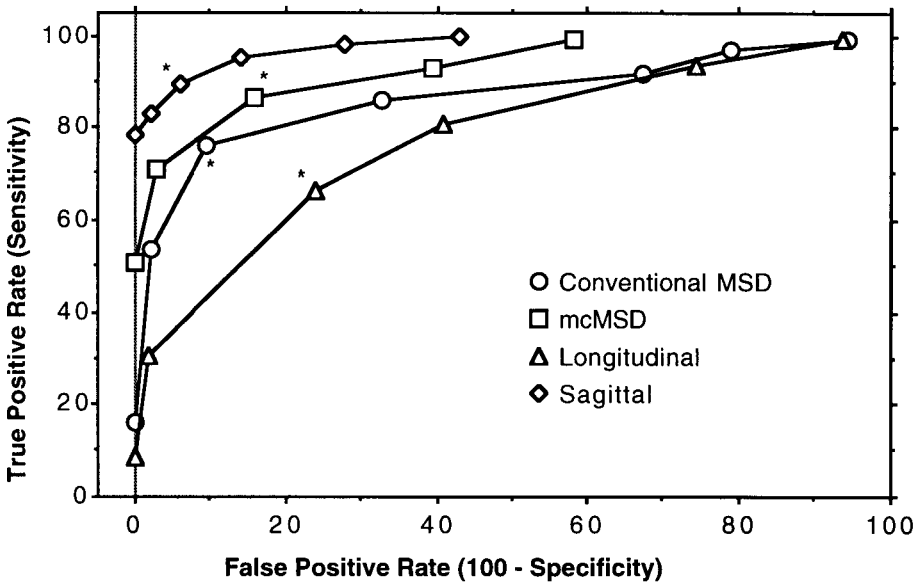


Figure 3.9 Response operating characteristic (ROC) curve for conventional and magnification-corrected MSD (mcMSD), sagittal and longitudinal ratio values at C6. The cutoff or operating points (MSD or ratio value) that maximize both sensitivity and specificity are identified with an asterisk. (Source of data: Moore, B.R., Reed, S.M., Biller, D.S., Kohn, C.W., and Weisbrode, S.E. 1994. Assessment of vertebral canal diameter and bony malformations of the cervical part of the spine in horses with cervical stenotic myelopathy. *Am. J. Vet. Res.* 55:5-13. With permission.)

also illustrates another important point: highly sensitive tests are most useful when their results are negative, and highly specific tests are most useful when their results are positive.

B. 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. By plotting ROC curves for each test, it is easy to determine which test provides the best criteria for ruling-in a disease over the range of each test's operating conditions.

EXAMPLE: Cervical stenotic myelopathy (CSM; also known as wobbler syndrome or equine sensory ataxia) is the leading cause of spinal ataxia of horses in most parts of the United States. Spinal ataxia results from spinal cord compression caused by malformation of the cervical vertebrae and narrowing of the vertebral canal, most frequently involving C5 - C7. The disease usually occurs within the first 1 - 2 years of life. Although definitive antemortem diagnosis requires myelographic examination, cervical radiographs, which permit calculation of the vertebral canal's minimum sagittal diameter (MSD, in millimeters), may be useful for screening patients. The value of survey radiography in CSM is still controversial, due in part to the effect of magnification in radiographs of standing horses. Moore et al (1994) compared three methods of vertebral canal diameter assessment that minimize (magnification-corrected MSD [mcMSD]) or eliminate (sagittal and longitudinal ratio) the effects of radiographic magnification with the conventional MSD method in CSM-affected and -unaffected horses. Disease status (gold standard) was established by a combination of myelography, histologic

examination of the spinal cord, and neurologic examination. ROC curves for vertebral sites C4 through C7 were generated to compare the ability of conventional MSD, mcMSD, sagittal ratio, and longitudinal ratio methods to discriminated between CSM-affected and -unaffected horses. To facilitate comparative and statistical analysis, each ROC curve was quantitatively assessed by calculating the area under the curve. The sagittal ratio method was the most accurate for distinguishing between CSM-affected and unaffected horses at vertebral sites C5, C6, C7, and overall. Figure 3.9 compares the results obtained at C6. It is clear that the sagittal ratio ROC curve has the highest sensitivity/specificity combination over its entire operating range. The cutoffs or operating points (MSD or ratio value) that maximize both sensitivity and specificity for each method are identified with an asterisk. The authors caution that the actual cutoff for establishing a diagnosis of CSM will depend, in part, on the consequences of misdiagnosis (for example, euthanasia versus follow-up myelography).

C. USE OF LIKELIHOOD RATIOS TO COMPARE INFORMATION CONTENT OF DIAGNOSTIC TESTS

The likelihood ratio can also be used to select the test with the greatest utility (information gain) for ruling in or out a particular disease during a diagnostic workup. The test that yields the highest value for the expression

$$\text{MAX}\ \{\text{abs}[\text{Ln}(\text{sens}/[1 - \text{spec}])], \text{abs}[\text{Ln}([1 - \text{sens}]/\text{spec})]\}$$

is the test of choice (Warner et al, 1988), where MAX is the maximum value obtained from either of the two expressions separated by the comma. This expression says, "given the sensitivity and specificity of all tests, the results of this test provide the strongest evidence for the presence or absence of the disease in question." The use of logarithms and absolute values facilitate direct comparison of likelihood ratios for positive and negative test results.

VI. SOURCES OF BIAS IN THE EVALUATION OF DIAGNOSTIC TESTS

A. RELATIVE VERSUS ABSOLUTE 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 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.

The argument could be made that the evaluation of an ELISA test for *M. paratuberculosis* infection in cattle described earlier in this chapter really measured relative versus absolute test performance, because the gold standard, fecal culture, is itself prone to error. However, the rigid criteria used by the authors to define absence of infection (negative herd history for 15 years and negative herd fecal culture, or absence of signs and negative results on at least three cultures) make it unlikely that misclassification significantly affected the results of the study.

B. 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

Table 3.5 Surgical diagnosis of 106 cattle in the control group with clinical findings consistent with traumatic reticuloperitonitis

	Diagnosis	*No. of Animals*
1.	Intussusception/Intestinal Obstruction	16
2.	Peritonitis	14
3.	Vagus Indigestion/Bloat	13
4.	Abscess	
	Abdomen	2
	Abomasum	1
	Liver	6
	Omasum	1
	Ruminoreticulum	2
5.	Abomasal Ulcers	11
6.	Johne's Disease	10
7.	Lymphosarcoma	3
8.	Fatty Liver	3
9.	Indigestion	2
10.	Distended Small Intestine	2
11.	Diarrhea	1
12.	Unknown	19
	Total	106

From Dubensky, R.A. and White, M.E. 1983. The sensitivity, specificity and predictive value of total plasma protein in the diagnosis of traumatic reticuloperitonitis. *Can. J. Comp. Med.* 47:241-244. With permission.

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 (Ransohoff and Feinstein, 1978).

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 challenge in the comparison group is to determine whether (and when) the test yields false-positive results. A distinction must be made between a screening test, which is intended to be used in a random, apparently healthy population, and a diagnostic test, which is intended to be used among animals showing similar clinical signs (Center et al, 1986). In the former, apparently healthy animals are used as the nondiseased group. In the latter, 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 (White, 1986).

EXAMPLE: Dubensky and White (1983) evaluated the use of total plasma protein (TPP) in the diagnosis of traumatic reticuloperitonitis (TRP) in 169 dairy cattle. Sixty-three cows with surgically confirmed TRP served as cases while 106 cows surgically explored for other abdom-

inal diseases that might be confused with TRP during differential diagnosis were controls. The presenting clinical signs in the two groups were similar and included anorexia, abdominal pain, bloat, colic, dehydration, depression, diarrhea, decreased milk production, fever, increased or decreased heart rate and weight loss. The surgical diagnoses (gold standard) for the control group included at least 12 distinct disease syndromes that should be on the differential list with TRP (Table 3.5).

A distinction must be made between a screening test, which is intended to be used in a random, apparently healthy population, and a diagnostic test, which is intended to be used among animals showing similar clinical signs.

C. 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). The first two occur when test results are available before a diagnosis is established.

Work-up bias occurs before a diagnosis is made and arises 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 after a diagnosis is made and arises when the result of a test affects 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.

EXAMPLE (of incorporation bias): Many of the hematologic and serum biochemical findings ranked by DiBartola et al (1987) (see Table 3.2) in cats diagnosed as suffering from chronic renal disease were, in fact, part of the case definition. These included azotemia, hyperphosphatemia, metabolic acidosis, mild hyperglycemia, lymphopenia, and nonregenerative anemia. This would tend to artificially increase their frequency among the diseased group. As a result, the case for their use in ruling out disease would be weakened.

The *disease spectrum*, e.g., the distribution of infection stages in the population used to evaluate the test, can also affect measurements of test sensitivity and specificity. For example, the sensitivity of an ELISA test for *M. paratuberculosis* infection of cattle is known to increase from approximately 25% to almost 90% as the disease progresses through the three successive stages of infection. Thus, estimates of test sensitivity may vary greatly depending on the age distribution and infection history of herds used to evaluate the test (Collins and Sockett, 1993).

VII. STATISTICAL SIGNIFICANCE

Often journal articles report that a diagnostic test was able to detect a "statistically significant difference" between control and infected groups. The magnitude of this difference may not be great enough to be clinically useful in the individual, however. 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.

VIII. SUMMARY

Diagnostic tests play a major role in the medical decision-making process. The gold standard refers to the means by which one can determine whether a disease is truly present or not. It provides a standard with which the performance characteristics of diagnostic tests can be evaluated.

Test sensitivity is defined as the proportion of infected or diseased individuals that test positive. Test specificity is the proportion of disease-free individuals with a negative test. 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. A sensitive test, i.e., one that is usually positive in the presence of disease, is frequently used to rule out certain diseases. A sensitive test is preferred when we don't want to miss a disease. A sensitive test may also be preferred as a screening test during the early stages of a diagnostic work-up, when a great many diseases are being considered, to reduce (rule-out) the number of possibilities. Specific tests, i.e., those that are rarely positive in the absence of disease, are useful to confirm, or rule-in, a diagnosis that has been suggested by other findings. They are especially useful when a false-positive diagnosis can result in physical, emotional or financial burden to the patient or owner. Thus, a sensitive test is most helpful to the clinician when the test result is negative, and a specific test is most useful when the test result is positive.

The probability of infection or disease in an individual with a given test result 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). Whereas sensitivity and specificity are absolute properties of a test, predictive value is relative, varying in response to changes not only in test sensitivity and specificity, but also the prevalence of disease in the population from which the patient came.

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. However, its value is subject to the same constraints as predictive value and is correct only for the population used to standardize the test. Reproducibility 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.

Clinical values in normal and diseased animal populations usually overlap, particularly when measured on an interval scale. Consequently, there is no way to adjust the positive/negative cutoff so that sensitivity and specificity are improved simultaneously. The optimal cutoff can be determined by selecting the cutoff yielding the lowest total number of incorrect diagnoses at a given pretest probability (or prevalence) of disease. Receiver-operating characteristic (ROC) curve analysis can also be used to identify the optimum cutoff. Both approaches can be made more clinically relevant by including the relative cost of false-positive and false-negative test results in the analysis. The likelihood ratio, ROC analysis and Bayes' graphs can also be used to compare the information content of diagnostic tests.

Three sources of bias in the interpretation of diagnostic tests are (1) improper standards of validity, (2) the spectrum of patients, and (3) prior knowledge of the health or disease status of individuals. An improper standard of validity is the use of an accepted standard test to calculate the relative sensitivity of a diagnostic test. Comparisons of the relative accuracy of one test over another are only valid when the true health status of test animals can be determined. Bias in the spectrum of patients occurs when the prevalence of the condition in the population to be tested differs from that used to standardize the test. Prior knowledge of the disease status of an individual may bias the time and effort expended to determine the sensitivity of a diagnostic test, thus increasing the likelihood that a positive test result will be reported. If we

were to examine the same samples without knowledge of the patient's clinical status, we might find the diagnostic test to be less sensitive.

Chapter 4

USE OF DIAGNOSTIC TESTS

I. INTRODUCTION

The value 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 (see Table 2.6). Indiscriminate testing at the herd level tends to reduce the predictive value of tests and can lead to unnecessary depopulation 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.

II. THE TESTING BAND

When managing a patient suspected of having a particular disease, three options typically exist: (1) perform other diagnostic tests to eliminate other possible differential diagnoses before instituting treatment; (2) institute treatment for the suspected disease regardless of test result; or (3) let test results guide treatment decisions (Smith, 1993). In part, this decision depends on the risks of the diagnostic procedure, the toxicity of therapy, and the difference between the outcomes for treated and untreated patients. In general, we withhold treatment when the probability of disease is low, treat empirically when the probability of disease is high, and let test results guide case management when the probability is intermediate. *But how low is low, and how high is high*?

The answer to the preceding question can be approached through a process called *decision analysis*. In deciding whether or not to perform a diagnostic test (or be guided by it), there are three strategic "regions" that are defined by two *threshold* possibilities: the *testing threshold* and the *treatment threshold* (Pauker and Kassirer, 1980). The probability of disease in a patient is compared with these thresholds. If that probability is below the testing threshold, then therapy should be withheld, even if the test is performed and the results are positive. If the probability of disease is between the testing and treatment thresholds (the "testing band"), then case management should be guided by the test result. If the probability of disease is above the treatment threshold, then empirical therapy should be given, even if test results are negative.

The two thresholds depend on five factors:

(1) the benefit of appropriate therapy = average gain in utility (compared with untreated individuals) among those who have the target disorder and receive appropriate therapy;
(2) the risk or cost of inappropriate therapy = average loss in utility (compared with appropriately treated individuals) among those who do not have the target disorder and are incorrectly treated;
(3) the risk of the test = average loss in utility associated with the testing procedure;
(4) the sensitivity of the test; and
(5) the specificity of the test.

These five factors are incorporated into two equations. The *testing threshold* equals

$$\frac{[(1 - \text{specificity}) \times (\text{risk of therapy})] + (\text{risk of the test})}{[(1 - \text{specificity}) \times (\text{risk of therapy})] + [(\text{sensitivity}) \times (\text{benefit of therapy})]}$$

The *treatment threshold* equals

$$\frac{[(\text{specificity}) \times (\text{risk of therapy})] - (\text{risk of the test})}{[(\text{specificity}) \times (\text{risk of therapy})] + [(1 - \text{sensitivity}) \times (\text{benefit of therapy})]}$$

To demonstrate how these equations are used, let us consider how much confidence should be placed in a serodiagnostic test for heartworm disease. In this example the risks and benefits are expressed in terms of likelihood of a favorable outcome.

EXAMPLE: Serodiagnostic tests for heartworm antigen are frequently used in the differential diagnosis of occult heartworm disease, e.g., patients showing signs consistent with the disease but whose blood is free of detectable microfilariae of *Dirofilaria immitis*. Let us consider the utility of one of the heartworm antigen detection kits over a broad range of pretest probabilities (prevalences) of infection (Smith, 1993).

Benefit of appropriate therapy: Approximately 20% of dogs suffering from heartworm disease will experience complications from thiacetarsamide treatment leading to death (utility = 0). Two-thirds (67%) of the remainder will become free of clinical signs. The likelihood of improvement for untreated heartworm disease is ≤5%. Thus, the benefit of appropriate therapy in heartworm disease = (.2 x 0) + (.8 x .67) - .05 = .486

Risk or cost of inappropriate therapy: The likelihood of improvement for correct and incorrect treatment of diseases other than heartworm were set at 67% and 5% respectively, as for heartworm disease. However, there is a chance (≤5%) of death associated with thiacetarsamide treatment of non-heartworm disease. Thus, the cost of inappropriate therapy in non-heartworm disease = .67 - (.05 x 0) - (.95 x .05) = .6225

Risk of the test (blood collection) = 0.

Sensitivity of the test = 65%, specificity of the test = 97.3% (Courtney et al, 1990).

The *testing threshold* equals

$$\frac{(.027 \times .6225) + 0}{(.027 \times .6225) + (.65 \times .486)} = .05$$

The *treatment threshold* equals

$$\frac{(.973 \times .6225) - 0}{(.973 \times .6225) + (.35 \times .486)} = .78$$

Thus for patients with pretest probabilities of occult heartworm disease below 0.05, it is best to avoid therapy and testing. For patients with pretest probabilities between 0.05 and 0.78, it is best to perform the test and treat with thiacetarsamide if the result is positive, and pursue other diagnoses if the test result is negative. Patients with pretest probabilities greater than 0.78 should receive thiacetarsamide, regardless of test result. In the author's experience, no more than one third of dogs from whom samples are obtained for occult heartworm testing are infected. Thus, routine testing is justified in this region of the country.

Recently the American Heartworm Society (1993) recommended that antigen tests replace the detection of circulating microfilariae as the accepted screening test prior to placing dogs on a preventative. Although the prevalence of heartworm infection may be less than 5% in this population of dogs, it does not mean that the use of antigen tests to screen dogs for heartworm disease should be discouraged. Under these conditions, the utility of treating or not treating diseases other than heartworm is ≈1.00, e.g., the prognosis that a healthy, uninfected dog will remain healthy, even if inappropriately treated with thiacetarsamide. Consequently, the risk of inappropriate therapy would be only 5%, e.g., the chance of an adverse reaction to heartworm therapy in an uninfected dog. Substituting this value in the above threshold analysis reduces the testing threshold from 0.05 to 0.004. As long as the prior probability for heartworm disease is ≥0.4%, as is the case in most heartworm-endemic regions, testing for heartworm disease is the best option. Under these conditions the treatment threshold drops from 0.78 to 0.22, reflecting the reduced risk of inappropriate therapy. If the benefit of appropriate therapy were to be increased, reflecting a more favorable prognosis in subclinically infected dogs, then both thresholds would drop even further.

In general, testing and treatment thresholds decrease as the risk of therapy decreases or as the benefit of therapy increases. Both thresholds increase as the risk of therapy increases or as the benefit of therapy decreases, i.e., one would need to be more certain that the disease is present to either test or treat. The testing band widens (the range of prior probabilities over which testing is appropriate) as the risk of testing decreases or as the sensitivity and specificity increase. The testing band narrows and may in fact disappear as the risk of testing increases or as the sensitivity and specificity decrease. Testing and treatment thresholds are further discussed in Chapter 14 in the context of decision analysis.

We can extend the threshold concept to include more than one disease state. In general, the ability to identify a second disease (which can be treated) with the same test will widen the testing band by allowing some patients who would have received either no treatment or empirical therapy for the wrong disease to be candidates for testing.

The logical selection and interpretation of diagnostic tests can help deal with the uncertainties inherent in the practice of medicine. By making explicit how uncertainty affects the meaning of clinical information, and by considering how the results of diagnostic tests will be used *before* those tests are ordered, clinicians should be able to improve the quality and the efficiency of the care they provide.

III. CALCULATION OF THE PROBABILITY OF DISEASE FROM TEST RESULTS

In the previous chapter Bayesian formulas were used to estimate the post-test probability of disease given a positive or negative test result, thereby providing a way to compare the accuracy of diagnostic tests. Although Bayesian formulas are useful for estimating post-test probabilities, they do not lend themselves to use on the clinic floor or in the field. Likelihood ratios provide an alternate approach to interpreting the significance of test results. Increasingly, reports of diagnostic test performance include the likelihood ratios for different levels of test results, thus making this approach even more useful.

A. CONVERSION BETWEEN THE PROBABILITY OF DISEASE AND THE ODDS OF DISEASE

Regardless of the scale used to report test results (positive/negative versus multiple levels of a test result), the conversion between the probability and odds of disease is the same. The basic mathematical relationship is represented by the equation:

Pretest Odds x Likelihood Ratio (LR) = Post-test Odds

Table 4.1 Use of the likelihood ratio to estimate the post-test probability of infection of cattle with *Mycobacterium bovis* for a positive or negative test 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 ≥ 0.35)	0.35	0.54	4.81	2.60	0.72
Negative test result (ELISA < 0.35)	0.35	0.54	0.32	0.17	0.15

Adapted from data in Figure 3.4. The pretest probability of infection = 35%, the prevalence of fecal culture-positive cattle in the test population.

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 is infected (or 0.20 in 1). 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 ÷ 0.8 = 0.25 (to 1). If the odds of heartworm infection would be 0.25 (to 1), then the probability of heartworm infection is 0.25 ÷ 1.25 = 0.20.

B. USE OF THE LIKELIHOOD RATIO TO CALCULATE POST-TEST PROBABILITIES

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 *M. paratuberculosis* infection in cattle (Chapter 3). At the optimal cutoff the likelihood ratio for a positive test (ELISA ≥ 0.35) was 4.81 and for a negative test (ELISA < 0.35) was 0.32. Table 4.1 depicts the results for a positive and negative test result, assuming a pretest probability of infection of 35% (prevalence of fecal culture-positive cattle in Figure 3.4).

Thus, a positive test result would increase the likelihood of *M. paratuberculosis* infection from 35% to 72%, but a negative test result would decrease the likelihood of infection to only 15%. Whenever possible it is best to express the likelihood ratio for each level of a test re-

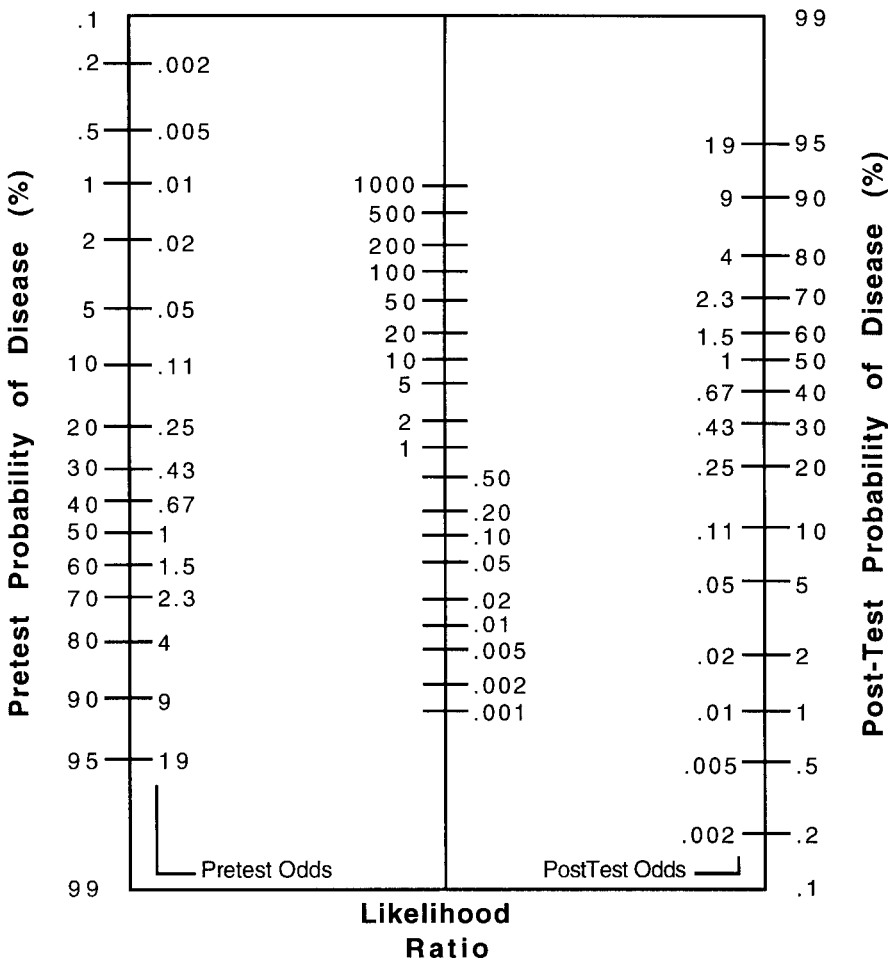


Figure 4.1 A nomogram for applying likelihood ratios and Bayes' theorem to the estimation of the post-test probability of disease. (Adapted from: Fagan, T.J. 1975. Nomogram for Bayes' theorem. *N. Engl. J. Med.* [letter] 293:257. With permission.)

sult, rather than above or below an arbitrary cutoff. For example, if the actual ELISA value were 0.85 (likelihood ratio = 26.40; Table 3.4), then the post-test odds of infection would be 26.40 x 0.54 = 14.26, increasing the post-test probability to 93%. This information is lost if test results are simply reported for just two levels, positive or negative.

C. 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 appears above (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 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 estimating the post-test probability of disease from the pretest probability for any level of a test result.

Consider, for example, the data in Table 4.1. To estimate the post-test probability of *M. paratuberculosis* infection for a positive test result, simply anchor a straight edge along the

Table 4.2 Aspects of multiple test strategies

	Test Strategy		
Considerations	*Parallel Testing*	*Serial Testing*	*Herd Retest*
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

left Y axis at a point approximating the 35% pretest probability of disease. Next, pivot the straight edge until it also lines up with a likelihood ratio of approximately 4.81. The straight edge should cross the right Y axis at about 72%, which is the post-test probability of disease. The same technique can be used to estimate the post-test probability of disease for a negative test result (likelihood ratio = 0.32), or for an ELISA value of 0.85 (likelihood ratio = 26.40).

D. 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). These concepts are expanded in subsequent chapters in which testing thresholds and decision analysis are discussed.

IV. MULTIPLE TESTS

Diagnoses are seldom made on the basis of a single test. Multiple testing is common in the veterinary hospital and in the herd. 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 4.2 summarizes the factors to be considered in ordering and interpreting multiple tests.

Table 4.3 Effect of parallel and serial testing on sensitivity, specificity and predictive value of test combinations

Test	*Sensitivity (%)*	*Specificity (%)*	*Positive Predictive Value (%)**	*Negative Predictive Value (%)**
A	80	60	33	92
B	90	90	69	92
A and B (parallel)	98	54	35	99
A and B (serial)	72	96	82	93

*For 20% prevalence.

From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology - The Essentials*, first edition, Diagnostic Strategies. Copyright 1982, The Williams & Wilkins Company. With permission.

A. PARALLEL TESTING

In parallel testing, two or more different tests are run on a patient or herd at the same time. 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 respiration are recorded. The degree of overlap in the distribution for 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. On the other hand, specificity and positive predictive value are lowered (Table 4.3). Only animals that have negative results on all tests are considered to be truly free of disease. The price is evaluation or treatment of some patients without disease.

EXAMPLE: Cats with hepatobiliary disorders often have vague clinical signs until the disease process is advanced (Center et al, 1986). High serum activities of certain liver enzymes often provide the first laboratory evidence of liver disease and may suggest the type of pathologic process developing in the liver. The diagnostic value of serum gamma glutamyl transferase (GGT) was compared with serum alkaline phosphatase (ALP) activity for the detection of liver disease in the cat. Sixty-nine cats (male = 36; female = 33) were examined because of suspected hepatic disease or because hepatic disease was considered after initial clinical observations. The diseased group consisted of 54 cats with histologically confirmed liver disease, while the control group consisted of 15 cats initially suspected of hepatic disease but subsequently found to be free of substantial histologic abnormalities of the liver. The study showed that GGT activity had superior sensitivity, but lower specificity, than ALP activity (Table

Table 4.4 Sensitivity, specificity and positive (pD+/T+) and negative (pD-/T-) predictive values of serum alkaline phosphatase and gamma glutamyl transferase activities when interpreted individually and in parallel in cats with liver disease

Test	*Sensitivity (%)*	*Specificity (%)*	*Positive Predictive Value (%)*	*Negative Predictive Value (%)*
ALP	50	93	96	36
GGT	86	67	90	59
ALP and GGT	94	67	91	77

From Center, S.A., Baldwin, B.H., Dillingham, S., Erb, H.N., and Tennant, B.C. 1986. Diagnostic value of serum gamma glutamyl transferase and alkaline phosphatase activities in hepatobiliary disease in the cat. *J.A.V.M.A.* 188:507-510. With permission.

4.4). The best sensitivity and negative predictive value (pD-/T-) was achieved using GGT and ALP activities simultaneously (in parallel). Therefore, the authors recommended that both tests be used in parallel to rule out the possibility of hepatobiliary disease in the cat.

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.

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. Referring to Table 2.6, if unrelated tests are performed in parallel, the chance that the 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. For the latter, serial or repeat testing is the preferred method.

B. SERIAL TESTING

In serial testing only those animals that tested positive on an initial test are retested. *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 the negative predictive value (see Table 4.3). We can be more confident in positive test results, but run an increased risk that disease will be missed. Because of the sequence in which serial testing is done, an animal is classified as affected with a condition only if it is positive on all tests.

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). Serial testing also decreases the likelihood of false-positive results due to time-related phenomena peculiar to particular patients. Examples are colostral antibody or vaccination titers. Along these lines, serial testing is frequently used in outbreak investigations in the form of paired sera collected at the time of a disease outbreak and two weeks later. A titer change to some specific pathogen provides presumptive evidence for its involvement in the outbreak.

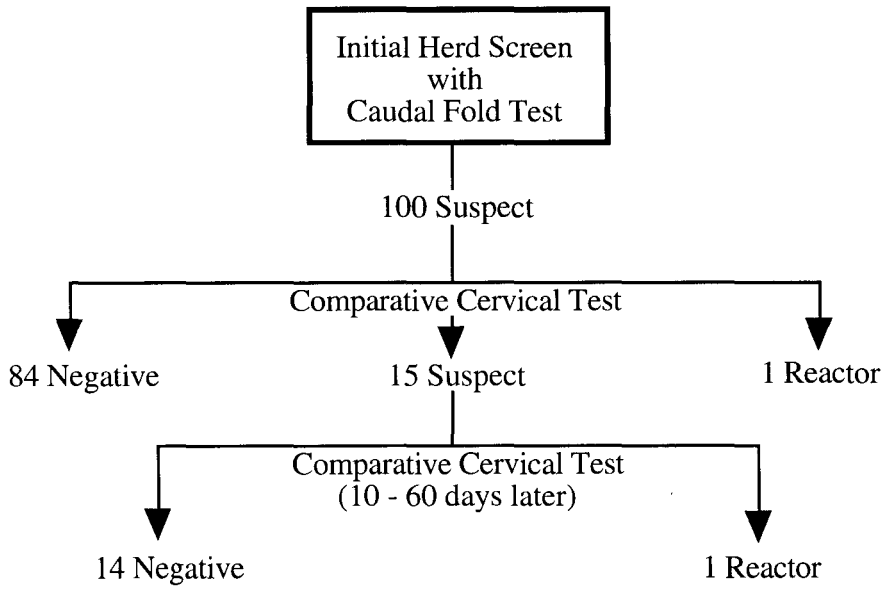


Figure 4.2 Recommended procedures for use of primary diagnostic and secondary supplemental tests for detection of reactors according to Uniform Methods and Rules of Bovine Tuberculosis Eradication Program. Typical outcomes for Illinois are included. (Source of data: McCoy [APHIS]. 1985. Personal communication.)

Serial testing is also an integral part of disease eradication 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. The Cooperative State-Federal Bovine Tuberculosis Eradication Program is an excellent example of serial testing. Accredited veterinarians play a key role in the program by applying the "primary diagnostic test," the caudal fold test, on a routine basis to herds. This is the official tuberculin test for routine screening of individual cattle, dairy goats and herds of such animals in which the tuberculosis status of the animals is unknown. The test measures the cellular reaction of cattle to the intradermal injection of purified protein derivative (PPD), which is extracted from *Mycobacterium bovis* organisms. Results are recorded as *negative* or *suspect*. Suspect animals are then retested by a state or federal regulatory veterinarian using the comparative cervical test (c-c test) to differentiate responses caused by mammalian tubercle bacilli and those induced by other mycobacteria. The c-c test in cattle is performed by injecting *M. avium* and *M. bovis* PPD tuberculins into separate sites in the skin of the neck. The difference in size of the two resultant responses usually indicates whether tuberculin sensitivity is caused by infection with bovine type bacilli rather than an avian type, by *M. paratuberculosis*, or by a transient sensitization to other saprophytic mycobacteria in the environment. These organisms are responsible for some of the false-positive tuberculin reactions that are a major problem in areas where tuberculosis has been nearly eliminated. Results may be negative, suspect or reactor (positive).

The proportion of reactors with no gross lesions can be greatly reduced by the use of the c-c tuberculin test. However, it is much more tedious, difficult to interpret and expensive than the caudal fold test. Consequently, it is not used as a primary screening test. To confirm a diagnosis of tuberculosis, it is necessary to isolate and identify the etiologic agent. Cultural results usually require six to eight weeks.

Table 4.5 Sensitivity, specificity and predictive values for an ELISA test for porcine trichinosis evaluated at two possible cutoffs

Parameter	*Cutoff 1 (%)*	*Cutoff 2 (%)*
Sensitivity	93.2	92.3
Specificity	92.3	95.6
Predictive values*		
Positive (pD+/T+)	93.2	96.3
Negative (pD-/T-)	92.3	90.9
Reported natural prevalence = 1/1000 (0.1%)		
Positive (pD+/T+)	1.2	2.1
Negative (pD-/T-)	99.9	99.9
Hypothetical prevalence = 1/10,000 (0.01%)		
Positive (pD+/T+)	0.12	0.21
Negative (pD-/T-)	99.9	99.9
Hypothetical prevalence = 1/100,000 (0.001%)		
Positive (pD+/T+)	0.01	0.02
Negative (pD-/T-)	99.9	99.9

*At the prevalence for this trial of 162/305 = 53%.

From Murrell, K.D., Anderson, W.R., Schad, G.A., Hanbury, R.D., Kazacos, K.R., Gamble, H.R., and Brown, J. 1986. Field evaluation of the enzyme-linked immunosorbent assay for swine trichinosis: efficacy of the excretory-secretory antigen. *Am. J. Vet. Res.* 47:1046-1049. With permission.

EXAMPLE: The typical results of retesting 100 animals classified as "suspect" by the caudal fold test are illustrated in Figure 4.2 (McCoy, 1985). The results of the serial testing sequence indicate a positive predictive value (D+/T+) for the primary screening test of approximately 2% (relative to the c-c test) and emphasize the importance of the secondary test in eliminating false-positive results from test results, especially when prevalence is low. Slaughter checks (the gold standard) have revealed that from 1% to 4% of Illinois cattle that are negative on two c-c tests actually contain visible lesions (McCoy, 1985). Thus, the negative predictive value (D-/T-) of serial testing is $\geq 96\%$. Once *M. bovis* is isolated from an animal, the herd can never be removed from quarantine and must be destroyed (McCoy, 1985).

EXAMPLE: Swine trichinosis provides another example of the need for confirmatory tests in disease control programs. The most recently reported national prevalence rate for swine trichinosis is 0.125% (Zimmerman and Zinter, 1971). Renewed interest in developing more efficient methods for the control of swine trichinosis has focused on the use of the enzyme-linked immunosorbent assay (ELISA) for use at both abattoir and farm. Murrell et al (1986) evaluated the efficacy of an excretory-secretory antigen in an ELISA screening test (Table 4.5). Although the positive predictive value of the test was greater than 90% in test herds, it would

decrease to 1% to 2% if applied to the U.S. swine population. The authors concluded that positive screening test results would have to be confirmed by follow-up serial testing with the more specific but tedious digestion procedure for an ELISA-based control program to be acceptable.

EXAMPLE: Expense is of prime concern in disease control or eradication programs, where parallel testing would result in unnecessary duplication of laboratory procedures, particularly as disease prevalence declines. The brucellosis milk ring test is used to screen all commercial dairies in the United States three to four times annually. Eradication efforts are concentrated in those herds that are identified as suspicious from the results of this test. A suspicious ring reaction is presumptive evidence of *Brucella* infection and is followed by a herd blood test. The milk ring test is remarkably specific; a rate of approximately 0.2% suspicious tests is now common in many states. Thus, more than 99% of the blood testing of dairy herds that would otherwise be necessary has been eliminated in those states (USDA, 1981).

It is perhaps comforting to know that if two tests are to be used in series, the number of test-positive and diseased individuals ultimately identified will be the same, regardless of the sequence in which the tests are used. *Everything else being equal, the test with the highest specificity should be used first, so as to reduce the number of individuals or samples that must be retested.*

Everything else being equal, when multiple tests are performed in series, the test with the highest specificity should be used first, so as to reduce the number of individuals or samples that must be retested.

C. HERD RETEST

Herd retest is a modification of serial testing with the exception that test-negative animals, 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 some time 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. Ongoing obligatory national programs include bovine brucellosis and tuberculosis eradication programs. Voluntary guidelines also exist for the eradication of such diverse diseases as bovine anaplasmosis and feline leukemia virus from their respective populations. In some cases the option exists to cull animals or treat them to remove infection, as is the case of bovine anaplasmosis.

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. We can envision a herd from which all infected individuals have been removed, i.e., prevalence of infection is 0%. If the test is only 80% specific, it will continue to report 20% false-positives (false-positive rate = 100 - specificity). If animals are methodically sent to slaughter after each round of testing, the herd may eventually be decimated.

D. ASSUMPTION OF INDEPENDENCE OF MULTIPLE TEST RESULTS

Ideally, when multiple tests are used, 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.

V. WORKING WITH DIFFERENTIAL LISTS

A. RULE-INS AND RULE-OUTS: THE CHOICE OF SENSITIVE OR SPECIFIC TESTS

The choice of a particular diagnostic test should be made within the context of the clinical situation. Negative results on a highly sensitive test, i.e., one that is usually positive in the presence of disease, is frequently used to rule out diseases on a differential list during the early stages of a diagnostic workup. Tests of high sensitivity are also useful when there is an important penalty for failure to detect a particular disease, or when the probability of disease is relatively low, as during the latter stages of disease eradication programs.

Sensitivity should not be the sole criteria for choosing a test, however. Use of a test of high sensitivity but low specificity in disease eradication programs employing test-and-slaughter may result in the destruction of an unacceptable number of animals with false-positive results and ultimately to depopulation of the herd. Positive results on highly specific tests, i.e., those that are rarely positive in the absence of disease, are useful to confirm (or rule in) a diagnosis that has been suggested by other data. They are especially useful when a false-positive diagnosis can result in physical, emotional or financial loss to the patient or owner.

B. INTEGRATION OF DIFFERENTIAL LISTS AND PROBABILISTIC REPORTING

The integration of a clinician's list of possible diagnoses with that of other specialists, such as pathologists, radiologists or nuclear medicine specialists, is a poorly defined process and is particularly prone to error. For example, a clinician receiving a biopsy report may not be fully aware that the pathologist has taken cognizance of the clinical facts in reaching a diagnosis. A clinician who assumes that a report reflects only the morphologic findings will be inclined to make the error of factoring in the clinical data a second time, i.e., "double counting" the clinical findings. On the other hand, without knowledge of a patient's clinical presentation, the pathologist may overvalue, i.e., give excessive weight to, the histologic findings.

Clinicians must interpret similar test results differently in different clinical situations. The interpretation of tests depends on the clinical situation and the presence of risk factors.

Schwartz et al (1981) have shown that a probabilistic approach to ranking diseases on a differential list reduces bias in estimates of diagnostic likelihoods. The approach calls for the clinician and the pathologist to evaluate their respective data independently. The clinician is asked to list the diseases under consideration and, before the biopsy, to provide a simple but numerical estimate of the probability that each disease would be present in a cohort of patients identical to the one in question (pretest probability of disease). The pathologist is asked to estimate the likelihood that the observed morphologic findings would occur in each of the diseases under consideration (test sensitivity). The two sets of figures are then used to calculate a revised estimate of the likelihood of each disease on the differential list (Table 4.6). Recalling Bayes' theorem for a positive test result (see Chapter 3), the clinician's estimate should be rec-

Table 4.6 Example of the probabilistic approach to pathology reporting using Bayes' theorem

Differential list of possible diagnoses	*A (Pretest Probability) Estimate of Likelihood of Each Disease Before Biopsy in a Cohort of Patients Identical to Sample Case*	*B (Test Sensitivity) Estimate of Likelihood of Observed Biopsy Findings in Patients with Each Disease*	*C Product of A x B*	*D (Post-Test Probability) Revised Estimate of the Likelihood of Each Disease on the Differential List (C x 100/Σ)*
Inflammation	65	5	325	25
Benign tumor	30	20	600	45
Malignant tumor	5	80	400	30
			Σ = 1325	

From Schwartz, W.B., Wolfe, H.J., and Pauker, S.G. 1981. Pathology and probabilities. A new approach to interpreting and reporting biopsies. *N. Engl. J. Med.* 305:917-923. With permission.

ognized as the pretest probability, the pathologist's estimate as the test sensitivity (sometimes referred to as the conditional probability), and the revised estimate as the post-test probability of each disease on the differential list.

The preceding example demonstrates the fact that clinicians must interpret similar test results differently in different clinical situations. The interpretation of tests depends on the clinical situation and the presence of risk factors.

VI. SCREENING FOR DISEASE

A. DEFINITIONS

When apparently healthy individuals are tested for the purpose of detecting disease, the process is referred to as *screening*. This testing strategy is called *case-finding* when applied to a clinician's own patients. Typically, every animal is sampled, and the objective is to identify the affected individual. Abnormal results are followed up with confirmatory diagnosis, then treatment or destruction of affected individuals.

When screening tests are applied to large, unselected populations this testing strategy is referred to as *mass screening*. Frequently, only a "statistically representative" sample of the herd or flock is sampled, and the objective is to identify affected populations rather than individuals. Identification of an affected population may then lead to case-finding through testing of each animal in the herd. In many official state and federal eradication programs, all animals are tested initially. Others, such as the swine pseudorabies eradication program, require testing of only a statistically representative number of animals. Case-finding and mass screening are extensively used in veterinary practice. A few examples are listed in Table 4.7.

Table 4.7 Some examples of case-finding and mass screening in veterinary practice

Case-Finding
A. In companion animal medicine
Routine history and physical examination
Routine coprological examination
Routine heartworm examination each spring
Feline Leukemia Virus testing
B. In herd health programs
Equine infectious anemia serologic testing
Bovine leukosis serologic testing
California mastitis testing of quarter milk
C. In regulatory medicine
Equine drug testing
Meat inspection
Mass Screening
Brucella ring testing of dairy herds
Market cattle testing of non-dairy herds
Caudal fold testing for bovine tuberculosis
Porcine pseudorabies serologic testing

B. 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 cheap, very safe and acceptable to both clients and practitioners.

VII. INCREASING THE PREDICTIVE VALUE OF DIAGNOSTIC TESTS

Considering the relationship between prevalence and 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 disease 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 operates in veterinary teaching hospitals, increases the likelihood of finding significant disease. Consequently, more aggressive use of diagnostic

Table 4.8 Criteria for abnormal physicochemical and microscopic findings in canine urine

Physicochemical	*Microscopic*
pH > 7.5	>10 RBC/hpf
Protein ≥ trace	>5 WBC/hpf
Glucose ≥ trace	>2 hyaline casts/lpf
Ketones trace	>1 granular cast/lpf
Occult blood trace	>1 waxy cast/lpf
Bilirubin > 2+	Microorganisms
	Parasitic ova/microfilariae
	Hyperplastic or neoplastic epithelial cells
	Unusual crystals

hpf = high-power (magnification) field (x450); lpf = low-power (magnification) field (x100).
Unremarkable crystals include triple phosphate, amorphous phosphate, calcium carbonate and bilirubin.

Reprinted with permission from Fettman, M.J. 1987. Evaluation of the usefulness of routine microscopy in canine urinalysis. *J.A.V.M.A.* 190:892-896.

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.

EXAMPLE: Historically, routine urinalysis has consisted of a parallel testing strategy combining macroscopic or physicochemical analysis and microscopic examination of the urine sediment (Table 4.8). Although microscopic examination increases laboratory-technician time and expense to the client, its use has been justified on the basis of high false-negative rates when physicochemical tests alone are used. Fettman (1987) explored the use of a risk-based testing strategy that reserved microscopic analysis for patients that were negative on physicochemical analysis but deemed at high risk for genitourinary disease.

The initial signs, clinical problems and results of urinalyses of 1000 consecutive canine patients examined at a veterinary medical teaching hospital were reviewed. Criteria for classification of high-risk patients included patient history, physical signs and clinical problems consistent with diseases in which genitourinary disease might be highly suspect (Table 4.9). Physicochemical examination alone would have incorrectly classified 64 of 562 individuals as

Table 4.9 False-negative rates (micropositive/ macronegative) for physicochemical tests of canine urine specimens by risk group

Diseases and Signs	*False-Negative Rate (%)*
Signs associated with predisposition to genitourinary disease	
Diarrhea[§]	60.0*
Perineal abnormalities[§¶]	45.5*
Genitourinary abnormalities[§†]	40.5*
Serum biochemical abnormalities due to renal failure[#]	37.5**
Neurologic deficits[¥∫]	35.3*
Lower vertebral disk abnormalities and hind limb lamenesses[∫]	29.2*
Signs not associated with predisposition to genitourinary disease	
Nasal neoplasia	15.8
Congestive heart disease	11.1
Neoplastic or inflammatory oral disease	10.5
Neoplastic or inflammatory lung disease	10.0
Skin neoplasia	7.1
No illness	3.5

[§]Clinical problems potentially associated with ascending genitourinary inflammatory disease.
[¶]Perianal neoplasia or inflammatory disease, perineal hernia.
[†]Dysuria, stranguria, pollakiuria, polyuria.
[#]High values for urea nitrogen, creatinine and/or phosphorus.
[¥]Caudal lower motor neuron or upper motor neuron deficits only associated with caudal muscular dysfunction.
[∫]Clinical problems potentially associated with impaired micturition.
*P = <0.001
**P < 0.08

Reprinted with permission from Fettman, M.J. 1987. Evaluation of the usefulness of routine microscopy in canine urinalysis. *J.A.V.M.A.* 190:892-896.

normal (micropositive/macronegative), resulting in a negative predictive value (D-/T-) of 88.6%. By performing follow-up microscopic urinalysis on 136 of the 562 macronegative individuals classified as being at high risk for genitourinary disease (Table 4.9), an additional 51 micropositive individuals would have been detected. As a result, only 13 of 426 macronegative individuals that tested negative would have been incorrectly classified, raising the negative predictive value of this multiple test strategy to 97.5%. This increase in negative predictive value was achieved despite the fact that 426 low-risk patients were not retested microscopically, thus resulting in considerable potential savings in laboratory-technician time and client costs. These savings would have to be balanced against the cost of failing to detect a patient with genitourinary disease.

VIII. SUMMARY

When managing a patient suspected of having a particular disease, three options typically exist: (1) perform other diagnostic tests to eliminate other possible differential diagnoses before instituting treatment; (2) institute treatment for the suspected disease regardless of test results; or (3) let test results guide treatment decisions. The range of prior probabilities of disease over which test results should guide case management defines the testing band. The lower and upper limits of the testing band are called the testing and treatment thresholds, respectively. The width of the testing band depends on five factors: (1) the benefit of therapy; (2) the risk or cost of therapy; (3) the risk of the test; (4) the sensitivity of the test; and (5) the specificity of the test. Decision analysis is a process whereby testing and treatment thresholds may be established.

The likelihood ratio can be used to estimate the post-test probability of disease for any test result. The results are the same as those derived using Bayesian formulas, but the likelihood ratio is easier to apply under clinical or field conditions. 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 the tests performed are conditionally independent, i.e., the results of one test are not related to those of another.

Multiple testing may be performed in three different ways: parallel, serial and herd retest. In parallel testing, two or more different tests are run on a patient or herd at the same time. In order to be considered free of disease, the patient must be negative on all tests. The net effect is to increase test sensitivity and negative predictive value, thereby increasing the probability that a disease will be detected. However, specificity and positive predictive value are lowered. In serial testing, only those animals that tested positive on an initial test are retested. An animal must be positive on all tests to be considered diseased. Serial testing maximizes specificity and positive predictive value, but lowers sensitivity and negative predictive value, thus increasing the probability that a disease will be missed. Herd retesting is commonly used in disease control and eradication programs. It is a modification of serial testing except that test-negative rather than test-positive animals are retested, usually after several months have passed. The net effect is to ask the herd to prove that it is free of the condition being sought by calling negative only those animals that are negative on all tests. Consequently, sensitivity (at the herd level) is increased. Use of tests of low specificity in disease eradication programs based on herd retesting may lead to excessive herd depopulation.

When apparently healthy individuals are tested for the purpose of detecting disease, the process is referred to as screening. This testing strategy is called case-finding when applied to a clinician's own patients. When screening tests are applied to large, unselected populations, this testing strategy is referred to as mass screening. Frequently, only a statistically represen-

tative sample of the herd or flock is sampled, and the objective is to identify affected populations rather than individuals.

The positive predictive value of tests can be improved by restricting testing to individuals at greatest risk of having the condition of interest. This can be accomplished through a referral process, by restricting testing to demographic groups (by age, sex, breed, etc.) known to be at greater risk of having the condition and by a carefully conducted screening history and physical examination before ordering additional diagnostic tests.

Chapter 5

MEASURING THE COMMONNESS OF DISEASE

I. 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 our current concepts of the risk of contracting a disease, 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.

II. EXPRESSING THE FREQUENCY OF CLINICAL EVENTS

A. 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*. 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...

$$\text{rate} = \frac{\text{numerator}}{\text{numerator} + \text{denominator}} \qquad \text{ratio} = \frac{\text{numerator}}{\text{denominator}}$$

and in the special case for disease...

$$\text{rate} = \frac{\text{affected}}{\text{affected} + \text{unaffected}} \qquad \text{ratio} = \frac{\text{affected}}{\text{unaffected}}$$

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: During the 1985-86 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. The ratio of female to male students was 1.034 to 1 (AVMA, 1986).

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

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: $\frac{\text{No. of affected individuals during an outbreak}}{\text{Population at risk at beginning of outbreak}} \times 10^x$	Useful for identifying risk factors for a specific disease; restricted to outbreak investigation
Incidence rate: $\frac{\text{No. of new cases of a disease over a time interval}}{\text{Average population at risk during time interval}} \times 10^x$	A dynamic measure of risk of acquiring disease over a given period of time; useful for monitoring the course of an epidemic; used in cohort studies to measure effect of suspected or known risk factors
Prevalence rate: $\frac{\text{No. of existing cases of a disease at a point in time}}{\text{Population at risk at same moment in time}} \times 10^x$	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

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.

From Schwabe, C.W. 1984. Analytical epidemiology and veterinary economics. *Veterinary Medicine and Human Health*, third edition. Williams & Wilkins, Baltimore, MD, pp. 430-447. With permission.

Table 5.2 Prevalence, listed by state, of *Mycobacterium paratuberculosis* isolated from the ileocecal lymph node in culled cattle

*State**	*No. Submitted*	*No. Infected Lymph Nodes*	*Prevalence (%)*	*Standard Error† (%)*
Alabama	106	1	0.9	0.9
Arkansas	102	1	1.0	1.0
California	531	8	1.5	0.5
Colorado	111	1	0.9	0.9
Georgia	104	0	0.0	1.2
Illinois	171	2	1.2	0.8
Kansas	394	0	0.0	0.6
Kentucky	129	1	0.8	0.8
Maine	101	1	1.0	1.0
Michigan	118	2	1.7	1.2
Minnesota	238	13	5.5§	1.5
Missouri	449	4	0.9	0.4
Mississippi	116	2	1.7	1.2
Nebraska	249	3	1.2	0.7
New York	365	5	1.4	0.6
Ohio	214	6	2.8	1.1
Oklahoma	415	1	0.2	0.2
Oregon	141	4	2.8	1.4
Pennsylvania	307	21	6.8§	1.4
Tennessee	197	4	2.0	1.0
Texas	1,215	9	0.7§	0.3
Virginia	193	3	1.6	0.9
Washington	236	2	0.9	0.6
Wisconsin	562	13	2.3	0.6
All others	776	12	1.6	0.4

*Only states from which at least 100 specimens were received are shown separately.

§States with prevalences that differed significantly from the overall mean, based on a Z test with a level of significance $P < 0.01$.

†Standard error is equal to the SD of the mean.

Reprinted with permission from Merkal, R.S., Whipple, D.L., Sacks, J.M., and Snyder, G.R. 1987. Prevalence of *Mycobacterium paratuberculosis* in ileocecal lymph nodes of cattle culled in the United States. *J.A.V.M.A.* 190:676-680.

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.

B. 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.

EXAMPLE: We have already seen some examples of prevalence in previous chapters. The rate of infection of U.S. swine with *Trichinella spiralis* (0.125%) discussed in the previous chapter is a prevalence rate. Another example is presented in Table 5.2, which lists the prevalence of *Mycobacterium paratuberculosis* infection, causative agent of paratuberculosis or Johne's disease, in cattle in selected U.S. states. The data were based on culture of ileocecal lymph nodes obtained from cattle culled at 76 slaughterhouses in 32 states and Puerto Rico during 1983 and 1984 (Merkal et al, 1987). The prevalence was 1.6% overall, with 2.9% in dairy culls and 0.8% in beef culls. The prevalence for female and male animals did not appear to differ significantly. We have no information on when the infections were acquired or the duration of infection. The rate thus represents the likelihood of being a case, rather than becoming a case. Standard error is defined as the standard deviation of the mean and is a measure of the variability of the reported prevalence values. See Chapter 2 for further information on the standard deviation. The derivation of standard error for this example is discussed in detail in Chapter 9 (Statistical Significance - Confidence Interval for a Rate or Proportion).

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

EXAMPLE: State testing laboratories receive many bat specimens from citizens, local agencies and veterinarians concerned about rabies. Although these samples of the wild population are biased, they can provide useful information about the dynamics of rabies in bats if carefully interpreted. Figures 5.1 and 5.2 depict yearly and monthly rabies prevalence rates in Illinois based on samples submitted to state agencies over an 18- to 22-year period (Burnett, 1989). The 5% to 6% overall prevalence estimate may be exaggerated because of sampling bias. Rabid bats frequently exhibit paralysis and are more likely to be captured than healthy bats. Furthermore, once rabid bats are identified, public concern is heightened and these bats are more likely to be captured and submitted for testing. Although significant differences in yearly prevalence occurred, suggesting long-term cycles of rabies in bat populations, the annual variation was relatively small (Figure 5.1). Mean monthly prevalence rates (Figure 5.2) indicate a marked seasonal pattern of winter lows and summer highs. The decline in early summer coincides with the influx of juveniles into the sample. These juveniles, which may not yet have developed rabies, are poor flyers and more likely to be captured. Although the actual incidence (new cases) of disease has not been measured, the repeated sampling strategy permits some generalizations about the likelihood of bats contracting rabies.

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

Incidence is the proportion of individuals 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 time period specified. It is, therefore, a measure of the risk of becoming a case over a defined time period. An example would be the monthly incidence of rabies in pet and wildlife populations. Since the rabies case fatality rate is 100% and occurs within a few days of the onset of clinical signs, we can be certain that each case is a *new case* over the monthly intervals. If rabies were a chronic dis-

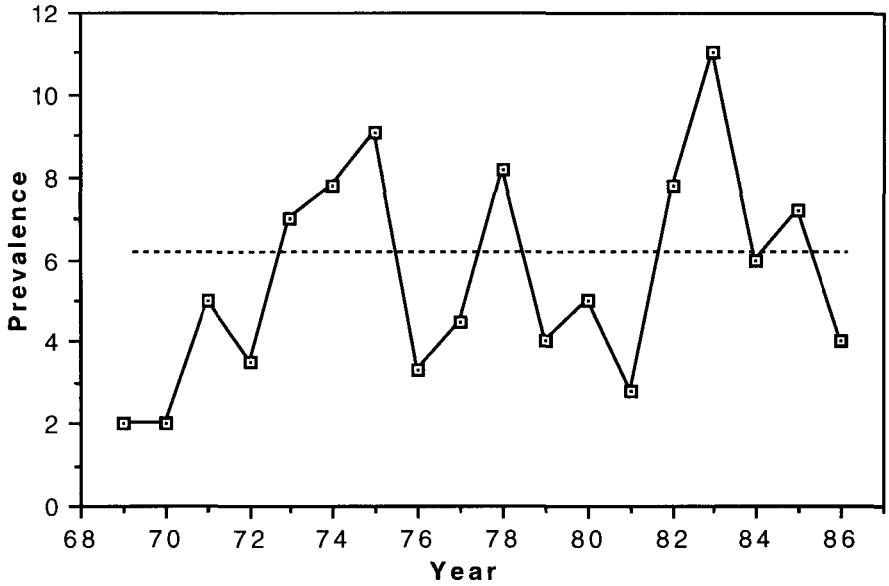


Figure 5.1 Annual rabies prevalence in 4272 bats submitted to the Illinois Department of Public Health over an 18-year period (1969 to 1986) by Illinois citizens and local agencies for rabies testing. The dashed line depicts the mean 18-year prevalence of 6%. (From Burnett, C.D. 1989. Bat rabies in Illinois: 1965-1986. *J. Wildl. Dis.* 25:10-19. With permission.)

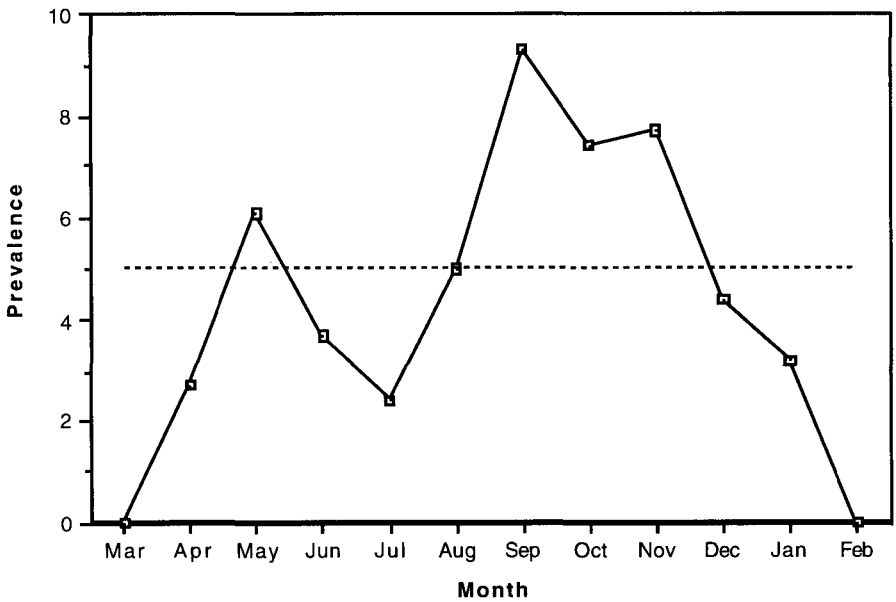


Figure 5.2 Monthly rabies prevalence in 2433 bats submitted to the Illinois Natural History Survey over a 22-year period (1965 to 1986) by Illinois citizens and local agencies for rabies testing. The dashed line depicts the mean 22-year prevalence of 5%. (From Burnett, C.D. 1989. Bat rabies in Illinois: 1965-1986. *J. Wildl. Dis.* 25:10-19. With permission.)

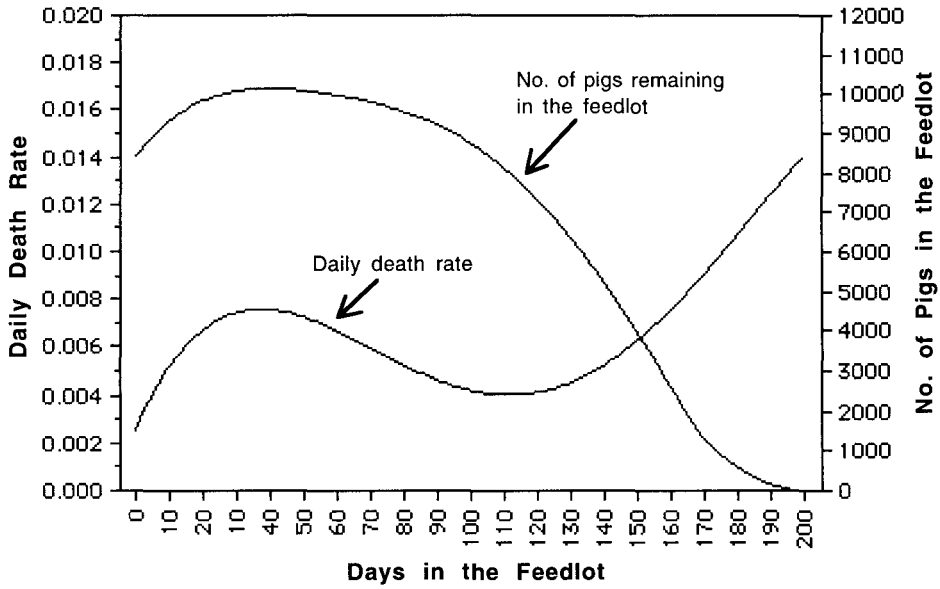


Figure 5.3 Comparison of daily death rate versus length of stay of swine in a feedlot. (From Straw, B.E., Henry, S.C., and Fleming, S.A. 1985. Interactions of management and animal performance in a swine feedlot. *J.A.V.M.A.* 186:986-988. With permission.)

ease it would be difficult to distinguish new cases from those that first appeared months earlier, such as occurs with many parasitic infections.

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 mean of the population at risk during that period (Figure 5.3).

EXAMPLE: Straw et al (1985) studied the effect of a number of management factors on productivity in a swine feedlot. One of the outcomes monitored was the daily death rate, which was calculated by dividing the number of deaths occurring each day (new cases) by the number of swine present in the feedlot at that time (population at risk). When adjusted for length of stay in the feedlot, it was found that the average incidence of death among pigs that failed to reach market weight within 150 days after entry into the feedlot (0.0104) was nearly twice that of pigs that reached market weight before 150 days (0.0054) (Figure 5.3). Although the specific cause(s) was not identified, the authors recommended that all animals be marketed by 150 days after entry into the feedlot, regardless of weight. The two lines cross at the 150-day mark. Note that a change in either of the y-axis (ordinal) scales would result in a shift of the crossover point and perhaps a different recommendation.

EXAMPLE: Figure 5.4 illustrates the difference between prevalence and incidence in a hypothetical population of 100 individuals studied over a 3-year period. Data are presented on the 18 individuals who contracted the disease. The other 82 remained uninfected over the three-year period. Since prevalence is measured at a single point in time, the prevalence of disease at the beginning of 1988 was 3/100, or 3%. The *prevalence of disease* was 8/100, or 8%, at the beginning of 1989. Two of the eight were actually included in the 1988 prevalence figure,

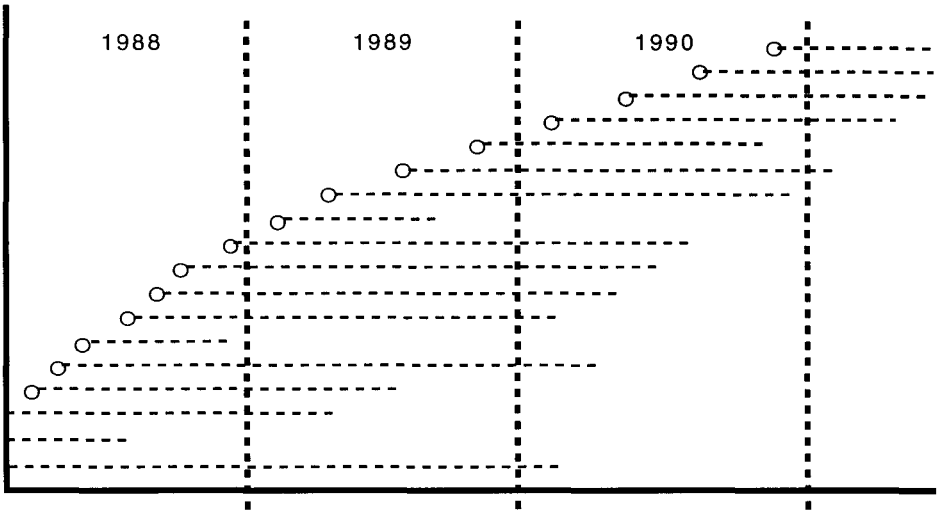


Figure 5.4 Occurrence of disease in 18 individuals from a hypothetical population of 100 animals over a 3-year period. (o) = disease onset; (- - -) = duration of disease signs, patency or serologic reactivity. (From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology - The Essentials*, first edition, Frequency. Copyright 1982, The Williams & Wilkins Company. With permission.)

but their disease persisted long enough to be detected in 1989. Prevalences at the beginning of 1990 and 1991 are 9% and 5%, respectively. Note that two animals, one in 1988 and one in 1989, contracted and recovered from disease between examinations and were thus lost or hidden from the prevalence survey.

The annual incidences of disease in the population depicted in Figure 5.4 would be calculated by dividing the number of new cases developing each year by the number of individuals free of disease at the beginning of each year. Since 3 of 100 individuals were already affected at the beginning of 1988, the annual incidences would be 7/97 (7.2%) in 1988, 4/90 (4.4%) in 1989 and 4/86 (4.7%) in 1990. In practice it is often impracticable or impossible to identify and exclude previously diseased individuals from the denominator in incidence studies. Under these conditions the population at risk for each yearly interval would be the entire herd, or 100 individuals.

Prevalence studies provide a variable index of the level of disease activity, depending on the manner in which disease is measured. This is due in part to the fact that the denominator may include animals that were previously infected and now immune. If prevalence was determined on the basis of the presence of clinical signs, then disease activity would appear to decline over time; not because the risk of contracting disease had diminished, but rather because fewer susceptible animals remained. On the other hand, if prevalence was determined on the basis of presence of specific antibody, disease activity would appear to be increasing over time because of the increased proportion of animals that had seroconverted.

EXAMPLE: The caprine arthritis-encephalitis virus (CAEV) is the causative agent of polyarthritis of adult goats or, rarely, leukoencephalomyelitis of kids. Exposure to the infective agent appears to be widespread in the United States. A serologic survey of sera from 24 states revealed serologic prevalence to be 81% in an agar-gel immunodiffusion test (Crawford and Adams, 1981). Many owners have expressed concern over the high prevalence of CAEV and its effects on productivity and longevity. Since the agent is known to be present in colostrum and milk, some have resorted to feeding heat-treated colostrum and pasteurized milk to kids in

Figure 5.5 CAEV serologic prevalence by age for goats reared by pasteurized and unpasteurized methods in 13 California dairies. Note that each vertical bar is cumulative, including animals that seroconverted during previous years. (From East, N.E., Rowe, J.D., Madewell, B.R., and Floyd, K. 1987. Serologic prevalence of caprine arthritis-encephalitis virus in California goat dairies. *J.A.V.M.A.* 190:182-186. With permission).

an attempt to reduce herd infection. To evaluate the effectiveness of pasteurization in reducing transmission, East et al (1987) compared serologic prevalence of CAEV infection in goats of different age groups in 13 California dairies (Figure 5.5). Although the prevalence of serologic reactors was higher in dairies that fed unpasteurized milk to kids, there was a direct association of increasing age with higher CAEV seroprevalence in both pasteurized and unpasteurized reared goats. This suggested that additional horizontal (contact) transmission of the virus, presumably during the milking procedure, was important in the spread of infection. It is important to note that increased serologic prevalence with age does not indicate that the greatest risk of infection occurs among higher age groups. Increasing serologic prevalence merely reflects the addition of newly infected animals to those already infected during previous years and still detectable serologically, e.g., serologic prevalence is cumulative. A rough estimate of incidence for each age group can be obtained by subtracting the serologic prevalence of the preceding age group.

Attack rate is a general term for the proportion of a defined population affected during an 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 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 useful for estimating the contribution of various risk factors to the onset or course of an epidemic and food-borne disease outbreaks in particular. Measurement of risk is discussed in greater detail in Chapter 6.

III. MEASURING THE FREQUENCY OF CLINICAL EVENTS

A. 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 prevalence or *cross-sectional studies.*

Prevalence studies can be based on the examination of a group of animals at a single point in time (*point prevalence*), on a single examination of each of a series of animals seen over a period of time (*period prevalence*) or a combination of the two. For example, determination of the proportion of swine afflicted with pneumonia during a slaughter check represents the proportion of cases in a population at a single point in time (point prevalence). The proportion of different types of neoplasia among all equine neoplasms diagnosed in an animal disease diagnostic laboratory over a 5-year period represents the cumulative results of individual diagnoses over time (period prevalence). Both are expressions of prevalence since we do not know when disease first appeared or how long it has lasted. Neither provides information on the risk of becoming a case (incidence), only on the risk of being a case (prevalence).

During the course of prevalence studies it is useful to also identify factors that might be associated with the presence or absence of disease.

EXAMPLE: The prevalence of infectious bovine keratoconjunctivitis (IBK) was estimated for cattle exposed to a variety of possible risk factors (Webber and Selby, 1981). Higher prevalences were associated with calves and yearlings, Hereford and Hereford-cross cattle, certain backgrounding cattle operations, herds with a history of IBK and vaccination of mature cattle against infectious bovine rhinotracheitis (IBR). Lower prevalences occurred in dairy and older cattle and were associated with the winter months. Surprisingly, fly control measures seemed to have no relationship to prevalence of IBK. Although these findings cannot be used to estimate the risk of cattle acquiring IBK, they do suggest measures for reducing the risk of disease.

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.

B. INCIDENCE

Incidence is measured by recording the appearances of a condition of interest over time in a population initially free of the condition. This study design, called a *cohort study* is discussed in detail in Chapter 6. Whereas time is assumed to be instantaneous in prevalence 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 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.

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 (as the pigs in Figure 5.3). This approach is useful to determine 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 or years at risk.

EXAMPLE: A four-year prospective study of the risk of developing canine testicular neoplasia incorporated cryptorchid (risk group) and matched controls into the study as they were identified by veterinary practitioners. The average follow-up period for all dogs was two years,

and ranged from one to four years for individual dogs. Since dogs were in the study for variable periods, incidence was calculated on the basis of dog years of observation. Dog years (the denominator) was estimated by dividing the total months of observation for all dogs in the risk group by 12 (Reif et al, 1979).

A disadvantage of the incidence density approach is the variable length of the follow-up period for members of the study. If long-term follow-up patients are systematically different from short-term follow-up patients, then the resulting incidence measures may be biased.

IV. FACTORS AFFECTING THE INTERPRETATION OF INCIDENCE AND PREVALENCE

A. TEMPORAL SEQUENCES

Prevalence studies can be used to obtain a static picture of a situation at a fixed point in time, e.g., 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 5.6.

EXAMPLE: The significance of temporal relationships can be appreciated if we return to the study of risk factors for IBK described earlier. Some of the factors, such as breed, sex and probably age, clearly preceded the clinical outcome. Others, such as vaccination history and fly control practices, may not have as we do not know whether the disease, IBK, preceded or followed exposure to the suspected risk factors. On the other hand, in the cohort study that examined the risk of testicular neoplasia in cryptorchid and normal dogs, the risk (cryptorchidism) clearly preceded the outcome (testicular neoplasia).

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

B. 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.

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

Prevalence of a disease in a population may be higher or lower than incidence, depending on the average duration of the disease. Diseases which are rapidly fatal, such as rabies, or of short duration, such as bovine mastitis, might have a higher incidence than prevalence. Chronic diseases, such as some parasitic infections, might be readily detected for long periods of time, and would be more likely to appear in a prevalence study.

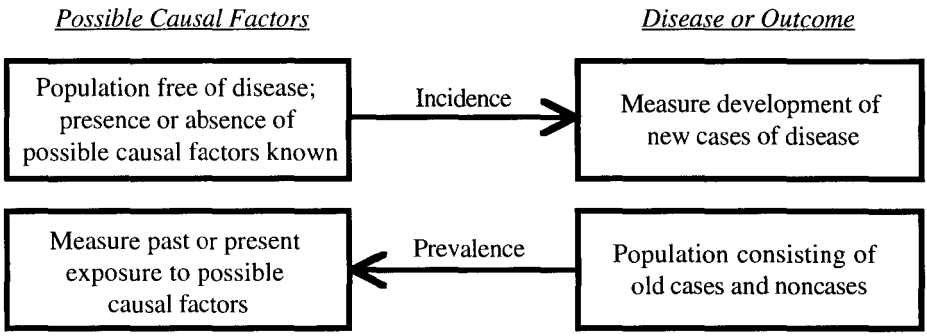


Figure 5.6 Temporal relationship between possible causal factors and disease: approaches based on incidence versus prevalence.

EXAMPLE: The routine surveillance reports of diseases present at slaughter published by the FSIS is biased in favor of diseases of long duration. For this reason the NAHMS was initiated. The NAHMS program uses federal, state and university veterinary medical officers to visit sampling units, or premises, where they collect incidence data through personal interviews, evaluation of herd health records and direct observations of the livestock or poultry (King, 1985). This approach provides better estimates of incidence than inspection-based surveillance systems.

C. 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 duration of disease (prevalence = incidence x average duration of disease). The equation can be rearranged to calculate any one of the parameters of interest.

EXAMPLE: Consider a dairy herd in which the prevalence of mastitis, based on the California Mastitis Test, is 4.5%. If the duration of a case of mastitis is typically three months (0.25 years) then the annual incidence of mastitis would be 4.5%/0.25 years or 18% of susceptibles per year. In other words, 18% of the herd will contract mastitis over the year, but at any given point in time only 4.5% of cows are diseased. The accuracy of this estimate of incidence depends in large part on the accuracy of our estimate of the duration of the disease.

D. REAL 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 real prevalence, which is usually estimated through use of an appropriate gold standard. If estimates of the sensitivity and specificity of a diagnostic test are available, it is possible to use them to estimate real prevalence 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 formula for estimating real prevalence from apparent prevalence is

Real Prevalence =

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

EXAMPLE: Collins et al (1994) conducted a random cross-sectional survey of Wisconsin dairy herds to determine the geographic distribution and prevalence of paratuberculosis, and to identify herd management factors associated with higher prevalence rates. An ELISA test with a sensitivity of 50.9% and specificity of 94.9% was used. Overall, 7.29% of cattle had positive test results. According to the equation

Real Prevalence =

$$\frac{7.29\% + 94.9\% - 100\%}{50.9\% + 94.9\% - 100\%} = 4.78\%$$

In this case, the real prevalence (4.78%) is a third less than the apparent prevalence (7.29%). This equation will not tell us whether a given test result is correct or not, but it does provide a better estimate of the true prevalence of disease.

E. 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 3.1) a list of clinicopathologic findings associated with chronic renal disease in cats was presented. The percentage of cats exhibiting any one finding ranged from 2.7% to 97%. As the number of signs required to diagnose chronic renal disease increases, the definition becomes more and more restrictive and includes a progressively smaller number of cases.

F. DANGLING NUMERATORS

Expressing the frequency of disease in terms of rates, using appropriate denominators rather than in terms of absolute numbers, e.g., *dangling numerators*, permits comparisons of disease rates 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).

EXAMPLE: *Cryptosporidium* sp. is a protozoan pathogen that can cause diarrhea in calves, lambs, goats, deer, immunocompromised as well as immunocompetent human beings and other domestic and wild animals. Sanford (1987) reviewed the records of all 3491 live pigs submitted to a veterinary diagnostic laboratory over a five-year period to obtain information on the characteristics of infected pigs. A total of 184 infected pigs from 133 farms were identified. He reported that the frequency of cryptosporidial infection was greatest in pigs 6 to 12 weeks old and that there was no seasonal pattern of infection. A perusal of the data from which these conclusions were drawn (Figures 5.7 and 5.8) reveals that his conclusions were based on the number rather than the rate of infection, i.e., dangling numerators. It is not clear whether the observed outcomes were real or whether the numerator data are merely a reflection of overall submission patterns. Had the data been expressed as rates, then the denominators would have been more comparable.

G. POPULATION AT RISK

Incidence and prevalence rates must be interpreted in the context of the population at risk.

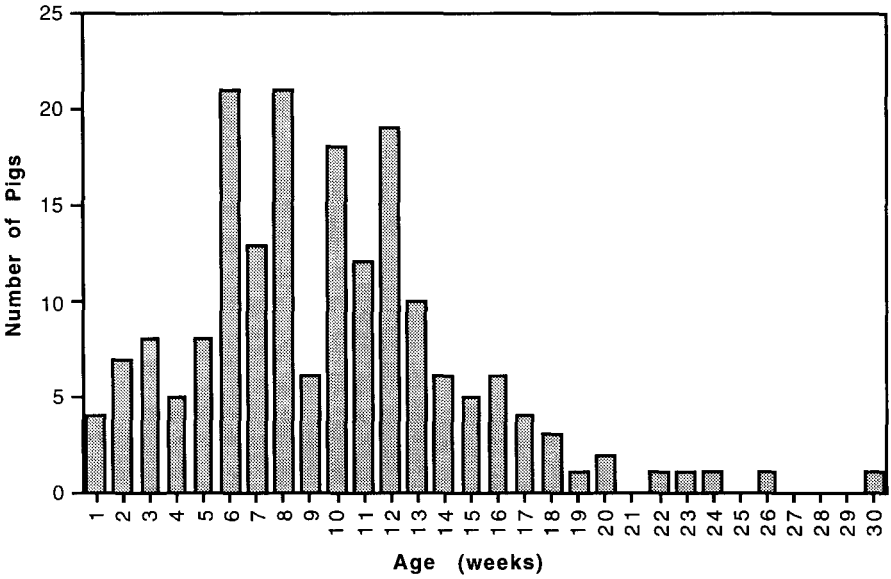


Figure 5.7 Age distribution of 184 pigs with enteric infection with *Cryptosporidium* sp. (From Sanford, S.E. 1987. Enteric cryptosporidial infection in pigs: 184 cases [1981-1985]. *J.A.V.M.A.* 190:695-698. With permission.)

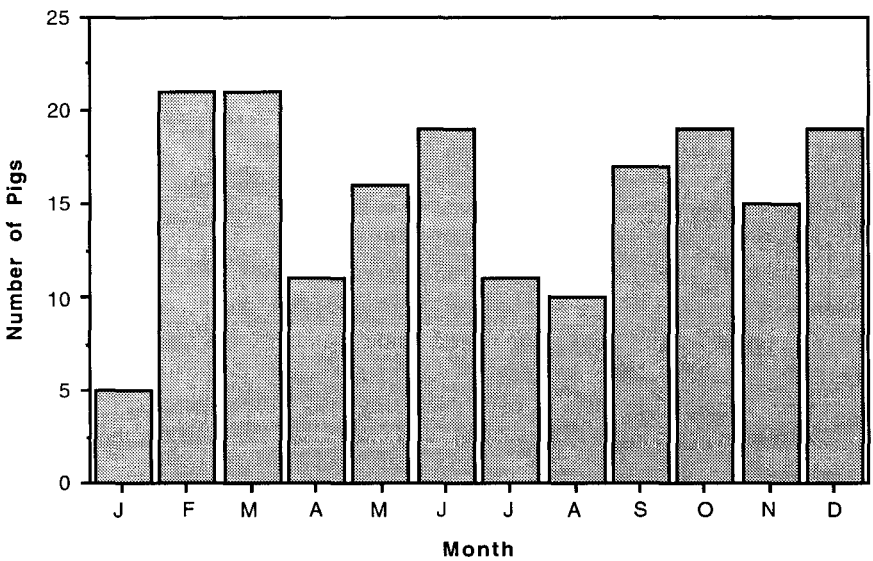


Figure 5.8 Monthly frequency of submissions for 184 pigs infected with *Cryptosporidium* sp. (From Sanford, S.E. 1987. Enteric cryptosporidial infection in pigs: 184 cases [1981-1985]. *J.A.V.M.A.* 190:695-698. With permission.)

If the population at risk differs significantly from one's own patients, then the rates may be meaningless.

Knowing which incidence or prevalence is representative of one's own patients is critical because of its influence on medical decision making. Because of the frequency of referral cases and usage patterns of veterinary services, the population of animals presented to the typical

veterinary medical teaching hospital (VMTH) is not representative of the population as a whole. This does not mean that the VMTH 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 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).

H. 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) *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 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 disease. Before morbidity or mortality rates in two populations can be compared, account must be taken of differences in age composition (Morton and Hebel, 1979). Consider the data in Table 5.3.

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 interest. When differences in the distribution of

Table 5.3 Death rates per 100 calves by age on two farms according to antibiotic use

	*Farm A Antibiotics Given to Calves**		*Farm B 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

*Antibiotics were being used therapeutically rather than prophylactically.

Source of age-specific death rates: Oxender, W.D., Newman, L.E., and Morrow, D.A. 1973. Factors influencing dairy calf mortality in Michigan. *J.A.V.M.A.* 162:458-460.

one or more host characteristics, such as age, occur among the groups we wish to compare, *adjusted rates* should be used (Morton and Hebel, 1979).

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.

V. ADJUSTED RATES: THE DIRECT METHOD

One method that can be used to adjust rates is referred to as the *direct method* (Kleinbaum and Kleinbaum, 1976). 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 *specific rates*. Each component of the sum (crude rate) has the following form:

Proportion of the population in each subgroup x 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 the two populations based on a *common* distribution for the confounding factor.

To compute direct 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 stand-in (substitute) for the different distributions of the two populations being compared.

A. 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 5.3. We arbitrarily define the standard population to be the sum of calves from the two

Table 5.4 Direct adjustment of death rates among calves on two farms according to antibiotic use

		Farm A Antibiotics Given to Calves		*Farm B 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, W.D., Newman, L.E., and Morrow, D.A. 1973. Factors influencing dairy calf mortality in Michigan. *J.A.V.M.A.* 162:458-460.

farms in each age group. The method for calculating age-adjusted death rates involves three steps and is presented in Table 5.4:

(1) Find the expected deaths for each age group by *multiplying* the standard population at risk by the test population rate for each age-specific group.
(2) Compute total expected deaths by *adding* expected deaths over all age-specific groups.
(3) Compute the direct rate by *dividing* the total expected deaths by the total standard population.

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. Antibiotics were not a contributing factor in the deaths of calves on Farm A. Rather, antibiotics were used *because of* the higher death rate and other disease problems on the farm.

B. 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 disease and death rates may be stated for the entire population or for any age, breed or sex subgroup.

EXAMPLE: Responses of atopic dogs to intradermal challenge with 60 allergens were determined and compared for four regions of the United States: northern Florida (n = 53), southern Florida (n = 67), Illinois (n = 130) and North Carolina (n = 28) (Schick and Fadok, 1986).

Responses to allergens were compared among the first three regions to determine their relative prevalence or frequency and whether significant ($P < 0.05$) differences existed, using Chi-square analysis. The number of patients seen in North Carolina ($n = 28$) was deemed too small for statistical analysis. Sex and breed prevalence of atopic dogs in northern and southern Florida were analyzed (Chi-square) for significant ($P < 0.05$) differences from the general hospital population at the University of Florida VMTH.

Sex and breed predispositions to atopy were detected. Females were found to have an increased tendency ($P < 0.05$) to develop clinical signs of atopy. West Highland White Terriers, Cairn Terriers, English and Irish Setters, Dalmatians, Lhasa Apsos and Golden and Labrador Retrievers were predisposed to develop atopy. Poodles had a significantly ($P < 0.05$) lower prevalence of atopy. Regional differences in responses to allergens were also found. Twenty-seven allergens incited significantly greater responses in dogs from northern Florida and 28 allergens in dogs from southern Florida, when compared with dogs from Illinois. Of Florida dogs with atopy, 79% had a positive response for flea antigen, compared with only 9% of dogs from Illinois. On the basis of these findings, the authors concluded that region-specific allergens should be used for diagnosis and hyposensitization treatment.

Though the previous findings are interesting and may be clinically useful, a nagging question is whether the results could be explained by the age-breed-sex composition of comparison groups. If age is a factor, could the breed and sex predisposition to atopy be the result of the age distribution of respective comparison groups? Likewise, could different age, breed or sex distributions in Illinois and Florida dogs explain the increased prevalence of atopy in Florida dogs? Use of adjusted rates would have strengthened the validity of this study.

EXAMPLE: PigCHAMP is a computerized record-keeping system for swine herds. It provides a valuable management and diagnostic tool for swine producers and veterinarians. One of the "outputs" of the program, the "Farm Comparison Report," compares a series of performance monitors, expressed as crude rates, for up to 12 farms. A number of these performance monitors, such as preweaning mortality, are parity specific, i.e., their values are known to be influenced by parity, e.g., number of litters the sow has produced (Stein and Duffy, 1988). Preweaning mortality rates exceeding 15% (action threshold) suggest that a problem exists that should be rectified. However, unless preweaning mortality rates are adjusted for parity, one may erroneously ascribe unacceptable mortality rates to disease rather than the age distribution of the sow herd.

In Table 5.5 actual crude preweaning mortality rates for two Illinois farms, Farm A (16.7% mortality) and Farm B (13.1% mortality) are adjusted by the direct method. Since preweaning mortality is calculated based on litters that are weaned or nursed off in the report period, the crude rates are adjusted for number of litters weaned for each parity group. The standard population chosen was that of Farm B, whose preweaning mortality was below the action threshold.

After rate adjustment we see that preweaning mortality for both farms is below the action threshold. A comparison of parity-specific mortality rates does not suggest overall difference between the two farms. The reason that the crude rates differed was due to a greater proportion of higher parity sows on Farm A (Figure 5.9), which generally have higher preweaning mortality values.

C. 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

Table 5.5 Direct adjustment of preweaning mortality rates on two Illinois swine farms according to parity of the sow; crude preweaning mortality rate for Farm A was 16.7% and for Farm B was 13.1%

		Farm A		*Farm B*	
Parity	*Standard Population at Risk*	*Preweaning Mortality (%) (p)*	*Expected Deaths (E)*	*Preweaning Mortality (%) (p)*	*Expected Deaths (E)*
1	133	13.6	18.1	13.2	17.6
2	130	13.7	17.8	9.7	12.6
3	120	2.5	3.0	9.8	11.8
4	105	10.4	10.9	15.0	15.8
5	80	20.0	16.0	16.5	13.2
6	45	21.6	9.7	16.6	7.4
7	30	19.7	5.9	21.0	6.3
8	39	37.5	14.6	11.5	4.5
Totals:	682		96.0		89.2

Direct rate (per 100) for Farm A $\frac{96.0}{682} = 14.1\%$

Direct rate (per 100) for Farm B $\frac{89.2}{682} = 13.1\%$

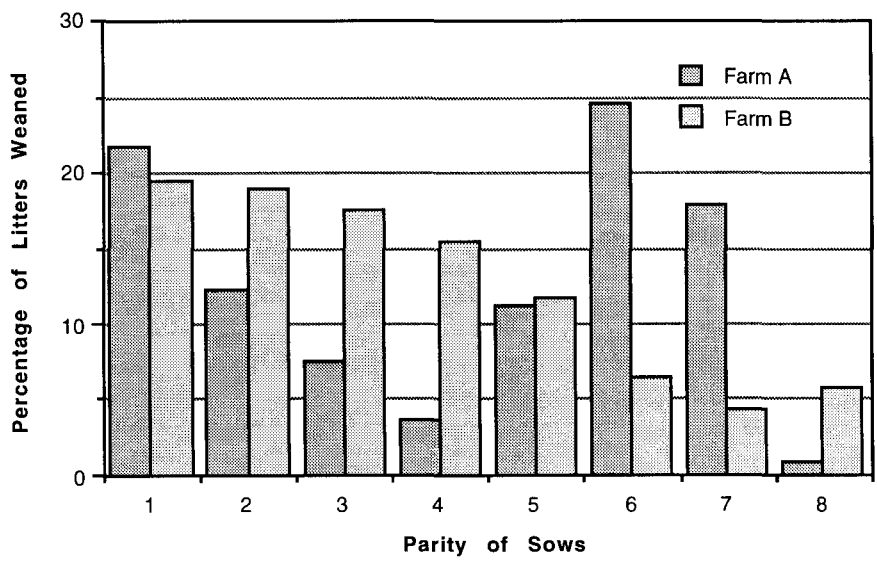


Figure 5.9 Comparison of parity distribution of litters weaned on two Illinois swine farms.

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 herd 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).

D. WHEN TO ADJUST RATES

Rates are adjusted in order to remove the effect of a factor that may confound a comparison. However, it is always necessary to first look at the overall crude rates, because they represent events. An adjusted rate gives an accurate comparison, but does not reveal the underlying raw data, which are shown by the crude rate (Morton and Hebel, 1979).

Although the presence of a (1) confounding factor is the primary condition for rate adjustment, three additional conditions must be met to justify adjusting rates:

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

Populations that appear comparable at first glance may in fact be found to differ in important ways if complete census data are examined. Adjustment of rates by age, sex or other relevant demographic factors may reveal differences that might otherwise be lost in the population as a whole.

E. 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.

VI. SUMMARY

Measurement of the frequency of clinical events is fundamental to current concepts of the risk of contracting a disease, its cause, prognosis and response to treatment. 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. 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 routinely deal with a number of rates. Some are vital statistics rates that can be used to 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. Prevalence is the proportion of sampled individuals possessing a condition of interest at a given point in time. 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. Incidence is the proportion of individuals that develops 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. Attack rate is a general term for the proportion of a defined population affected during an 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 of the outbreak. Since the attack rate is based only on new cases of the disease, it is comparable to incidence.

Sources of bias in prevalence studies include interpretation of the time sequence of suspected cause-effect relationships, inclusion of old as well as new cases and real versus apparent prevalence. The interpretation of incidence and prevalence rates also depends upon the degree to which cases and the population at risk are comparable to the populations that we are interested in. When making comparisons, rate adjustment is used to remove the effect of confounding factors, such as age, breed, and sex distribution, upon overall crude rates. The direct method of rate adjustment forces a comparison of populations based on a common distribution of the confounding factor.

Chapter 6

RISK ASSESSMENT AND PREVENTION

I. 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. In the previous chapter on frequency only rates were discussed. In this chapter ratios as well as rates are used 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 who is exposed. 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.

II. 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 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.

Table 6.1 Comparison of production indices in "average" and respiratory disease-free hogs

Disease Status	*Time to Market Wt. (days)*	*Average Daily Gain (lb)*	*Feed Conversion (lb feed/ lb gain)*
Respiratory disease free	170	1.47	3.4
Today's "average" hog	180	1.38	3.5

Source of data: Anonymous. 1985b. Study finds hog respiratory disease costing producers $200 million yearly. *DVM Magazine.* October, 1985, p. 57.

(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. This effect can be appreciated when the performance of respiratory disease-free swine is compared with their "normal" counterparts (Table 6.1).

(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, it has been claimed that 20% of a small animal practitioner's time is spent diagnosing or treating canine genetic diseases (Padgett, 1985). The risk of occurrence of genetic diseases in any particular individual, however, 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 is slow and involves large numbers of individuals to prove an association.

EXAMPLE: Ruble and Hird, (1993) examined 1679 6- to 18-week old dogs for congenital abnormalities over a 2-year period. Fifteen percent had at least 1 congenital defect and 1.5% had multiple congenital abnormalities. Defects observed, in descending order, were patellar luxation (7.2%), palpebral abnormalities (3%), cryptorchidism (2.6%), inguinal hernia (1.3%), faciodental malformations (1.3%), cardiac abnormalities characterized by murmurs (0.7%), and umbilical hernia (0.6%). Although practitioners are likely to encounter many cases of congenital abnormalities among their patients, detection of any breed associations would require systematic examination and record-keeping of a large number of such cases.

(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 metritis, mastitis, agalactia syndrome. For these diseases no single

cause can be identified. Rather, a combination of factors acting synergistically appears to be responsible for the disease syndrome.

III. 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. For example, in Table 5.3 antibiotic usage was associated with higher age-specific death rates among calves. However, antibiotics were being used to combat disease and did not cause death losses. Curtailing antibiotic usage would not have reduced calf losses.

(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 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 Chapter 10.

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.

IV. COHORT (PROSPECTIVE) STUDIES OF RISK

A. TRUE COHORT STUDY DESIGNS

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. 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 6.2. Examples of concurrent and historical cohort studies follow. An example of the use of both historical and concurrent cohorts in a clinical trial appears in Chapter 8 (Figure 8.2).

Table 6.2 Cohorts and their uses

Characteristic in Common	*To Assess Effect of*	*Example*
Age	Age (duration of exposure)	Effect of duration of cryptorchidism on incidence of testicular neoplasia (Reif et al, 1979; Chapter 5)
Date of birth	Calendar time	Effect of improved radiation safety procedures on incidence of lymphatic and hematopoietic tumors in veterinary practitioners (Blair and Hayes, 1982; Chapter 6)
Exposure	Etiologic agent	Effect of infection with feline leukemia virus upon mortality from selected diseases (Hardy et al, 1976; Chapter 7)
Disease	Prognosis	Prognosis for untreated feline dilated cardiomyopathy (Pion et al, 1992; Chapter 8)
Treatment	Therapeutic intervention	Prognosis for taurine-treated feline dilated cardiomyopathy (Pion et al, 1992; Chapter 8)

Table 6.3 A concurrent cohort study of risk in neonatal calves with various levels of serum gamma globulin

Gamma Globulin (%)	*Cohort Size*	*Deaths or Culls*	*Incidence (%)*	*Relative Risk**	*Attributable Risk**
1.1-6.2	73	12	16.44	12.16	15.09
6.3-12.0	73	3	4.11	3.04	2.76
12.1-19.3	73	2	2.74	2.03	1.39
19.4-46.7	74	1	1.35	1.00	0.00
Totals	293	18	6.14		

Adapted from data in Table 2.7.
*Compared with the high gamma globulin group.

Table 6.4 A historical cohort study of the risks associated with being a veterinarian (based on cause of death in 5016 white men, 1947-77)

Cause of Death	*Incidence in Veterinarians (%)*	*Incidence in General Population (%)*	*Relative Risk*	*Attributable Risk (%)*
All cancers (including the following)	16.59	16.39	1.01	0.20
Brain and CNS	0.56	0.34	1.63	0.22
Skin	0.48	0.30	1.61	0.18
Lymphatic and hemopoietic	2.23	1.50	1.49	0.73
Colon	2.21	1.65	1.34	0.57
Stomach	0.94	1.44	0.65	-0.51
Lung	2.29	3.71	0.62	-1.42
Suicide	2.73	1.60	1.70	1.13
Motor vehicle accidents	3.15	2.19	1.44	0.96
Circulatory disease	50.36	48.57	1.04	1.79
Respiratory disease	3.27	5.17	0.63	-1.90
All others	23.90	26.08	0.92	-2.18
Total	100.00	100.00		

From Blair, A. and Hayes, H.M., Jr. 1982. Mortality patterns among U.S. veterinarians, 1947-1977: an expanded study. *Int. J. Epidemiol.* 11:391-397. With permission.

1. Concurrent Cohort Studies

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

EXAMPLE: Let us return to our earlier discussion of the effect of low serum gamma globulin levels in newborn calves on subsequent survival (see Table 2.7). The data has been rearranged in Table 6.3 in a format that allows calculation of risk. Since the four groups of calves were assembled at one time, each group may be treated as a cohort and the outcome (survival or removal from the cohort) as incidence. Notice that the "Loss" column is now referred to as "Incidence" and that two additional parameters, designated relative and attributable risk, have been calculated. These are discussed in the following sections.

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.

EXAMPLE: Causes of death among 5016 white male veterinarians identified from obituary listings in the *Journal of the American Veterinary Medical Association* were compared with a distribution based on the general U.S. population, matched by 5-year age and calendar period (age-adjusted mortality; Table 6.4). Proportions of deaths were significantly elevated for can-

Table 6.5 A survival cohort study of the benefits of chemotherapy for advanced mammary adenocarcinoma in cats

Patient	*Breed*	*Age (yr)*	*OVH* (yr)*	*Duration of Signs (mo)*	*No. of Recurrences*	*Metastases to Thorax*	*Survival Time (d)*
1	DSH	9	1	14	2	No	NA
2	DSH	13	11	72	2	No	NA
3	Persian	11	7	3	0	Yes	NA
4	Siamese	13	11	24	3	Yes	4
5	Siamese	9	5	10	1	Yes	45
6	Siamese	11	9	24	2	Unknown	47
7	DSH	8	NA	8	0	Yes	67
8	Siamese	12	Intact	5	2	Yes	106
9	DSH	13	2	17	1	Yes	149
10	DSH	12	Intact	6	1	No	170
11	DSH	11	NA	9	2	Yes	180
12	DSH	14	Intact	16	1	No	182
13	Siamese	7	6	6	2	Yes	283
14	DSH	11	10	12	3	Yes	344

*Years since ovariohysterectomy.
DSH = domestic shorthair; NA = not available.

From Jeglum, K.A., de Guzman, E., and Young, K.M. 1985. Chemotherapy of advanced mammary adenocarcinoma in 14 cats. *J.A.V.M.A.* 187:157-160. With permission.

cers of the lymphatic and hematopoietic system, colon, brain and skin. Fewer deaths were observed than expected for cancers of the stomach and lung. Sunlight exposure was suspected for the excess of skin cancer among veterinarians whose practices were not limited exclusively to small animals. Ionizing radiation exposure was suspected for the excess of leukemia among veterinarians practicing during years when diagnostic radiology became widely used. Mortality was also high for motor vehicle accidents and suicides, but low for diseases of the respiratory system (Blair and Hayes, 1982).

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.

B. SURVIVAL COHORTS

Concurrent and historical cohorts are sometimes referred to as true cohorts, since they are studied from the point at which they 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 assembly of individuals is referred to as a *survival cohort.* The name does not imply that survival is being studied, but rather that each individual has survived, or is available for study, after a given period of exposure or disease.

Table 6.6 Advantages and disadvantages of cohort studies

Advantages	*Disadvantages*
The only way of establishing incidence (e.g., absolute risk) directly	Inefficient and expensive because many more subjects are included than experience the event of interest. Therefore, inappropriate for diseases of low incidence
Follow the same logic as the clinical question: if the subjects are exposed do they get the disease?	Can assess the effects of exposure to relatively few factors (i.e., those recorded at the outset)
Can assess the relationship between exposure and many possible outcomes (diseases)	Results not available for a long time

From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology - The Essentials*, first edition, Risk. Copyright 1982, The Williams & Wilkins Company. With permission.

EXAMPLE: Table 6.5 summarizes the results of a study in which a new chemotherapeutic regimen for advanced feline mammary adenocarcinoma was evaluated. This is a classic survival cohort in that the only thing the patients have in common is the particular type of tumor. The extent of tumor development among the patients when they were included in the study is highly variable. Aside from being in different stages of the disease, additional variables such as breed, age and ovariohysterectomy exist within the group. Survival time was measured from the start of chemotherapy to death (Jeglum et al, 1985).

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.

C. LIMITATIONS OF COHORT STUDIES

Some of the advantages and disadvantages of cohort studies are compared in Table 6.6. Since they are conducted in the present, concurrent cohort studies permit the collection of any data required for the specific purposes of the study. In contrast, data for historical cohort studies is 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, the study examining the risk of being a veterinarian (Table 6.4) 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. An example is the study of survival cohorts.

One of the major difficulties in cohort studies is assembly of all members of the cohort at the same time. As described in Chapter 5 (canine testicular neoplasia in cryptorchid dogs,

Reif et al, 1979), individuals exposed to a risk factor 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. If the outcome is infrequent, a large number of subjects must enter and remain in the study for a long time before results are available.

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 to assure that a sufficient number of cases will arise may make a cohort study impractical. An alternate approach, the *retrospective study*, is discussed later in this chapter.

D. COMPARING RISKS IN COHORT STUDIES

Incidence is the basic expression of risk. It is the number of new events (usually disease) arising in a defined population over a given period. Incidence is especially useful for evaluating the relationship between presumed risk factors and disease. Several measures, called *measures of effect*, are commonly used.

1. Relative Risk

Relative risk, 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.

2. Attributable Risk

Attributable risk, 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. If all cases are associated with the risk factor being measured, then attributable risk would be equal to the incidence of disease in the population as a whole.

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

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 herd, and which risks are associated with the greatest economic loss.

Population attributable risk is estimated by multiplying the attributable risk by the prevalence of the 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 disease incidence in a population than a stronger risk factor that is rarely present.

Table 6.7 Measures of effect in studies of risk of disease

Expression	*Clinical Question*	*Calculation**
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 x P
Population attributable fraction	What fraction of disease in a population is attributable to exposure to a risk factor?	AFp = ARp ÷ RT

*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.

From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology - The Essentials*, first edition, Risk. Copyright 1982, The Williams & Wilkins Company. With permission.

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 and is estimated by dividing the population attributable risk by the total incidence of 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 6.7 compares the various measures of effect for the risk of disease, while the following example describes how the indices can be used to describe the risks associated with low gamma globulin levels in neonatal calves.

EXAMPLE: As you will recall from Chapter 2, one way of defining abnormality is association with disease. The strength of this association is illustrated in Table 6.8 in which we continue to explore the impact of low serum gamma globulin levels on calf survival. For this analysis the lowest serum gamma globulin group is considered to be exposed to the risk factor, while members of the other three groups are pooled as controls. This approach would seem to be appropriate since the former group suffered by far the greatest calf losses, either through death or culling.

From the preceding analysis we can conclude the following:

1. Calves with low serum gamma globulin levels are approximately six times more likely to be culled or die than their "normal" counterparts (relative risk).

Table 6.8 Calculation of measures of effect: Suboptimal gamma globulin levels in calves*

Simple risks	
Incidence of calf losses among low gamma globulin group	= 16.44%
Incidence of calf losses among remaining calves	= 2.73%
Prevalence of low gamma globulin levels in all calves	= 24.91%
Incidence of calf losses	= 6.14%
Compared risks	
Relative risk	= 16.44 ÷ 2.73 = 6.03
Attributable risk	= 16.44 - 2.73 = 13.71%
Population attributable risk	= 13.71 x 24.91 = 3.42%
Population attributable fraction	= 3.42 ÷ 6.14 = 55.61%

Adapted from data in Tables 2.7 and 6.3.
*Low serum gamma globulin = 1.1% - 6.2%

2. Low serum gamma globulin levels are associated with an additional 13.71% incidence of culls and deaths among exposed calves (attributed risk).

3. Low serum gamma globulin levels are associated with an additional 3.42% incidence of culls and deaths among all calves (i.e., the herd, population attributable risk).

4. Low serum gamma globulin levels are associated with approximately 56% of calf losses among all calves (population attributable fraction).

V. CASE CONTROL (RETROSPECTIVE) 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 determine the proportion of "cases" and "non-cases" that were exposed to the factor(s) of interest? This approach, known as a *case control*

or *retrospective 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.

A. ADVANTAGES OF CASE CONTROL STUDIES

Case control studies lend themselves to clinical research since 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, in order to gather information about the risk factors for tuberculosis in 100 swine (current prevalence approximately 0.006%), a cohort of at least 1.5 million animals would have to be formed and followed from birth to slaughter. Obviously, the expense and logistic difficulties of such a study design would render it unrealistic. In contrast, it would be relatively inexpensive and easy to assemble 100 or more cases of swine tuberculosis, find similar groups of animals without the disease and compare frequencies of hypothesized risk factors.

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 the 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.

B. COHORT VERSUS CASE CONTROL APPROACHES

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 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.

To better appreciate the methodological differences between cohort and case control studies and potential sources of bias, consider how the two would examine the role of vaccination against infectious bovine rhinotracheitis (IBR) in the development of infectious bovine keratoconjunctivitis (IBK). A study (Webber and Selby, 1981) has suggested that such a relationship exists.

In the cohort approach sampling is based on exposure whereas in the case control approach sampling is based on outcome.

1. A Cohort Study of IBK Following IBR Vaccination

A cohort study designed to see whether vaccination against IBR predisposes beef cattle to IBK would begin by identifying a producer or group of producers with enough cattle to ensure that a sufficient number of cases would be seen (Figure 6.1). Each producer would be asked to provide data on the age, breeds, herd size, vaccination program, prior history of IBK and other management procedures. Each animal would be examined and those showing evidence of IBK or other disease at the time of initial screening would be removed from consideration. The remainder would be stratified (grouped) according to the previously mentioned factors, and a sample of cattle from strata would be selected randomly for inclusion into one of two cohorts – IBR vaccinated or unvaccinated. The cohorts would be monitored at regular intervals for ev-

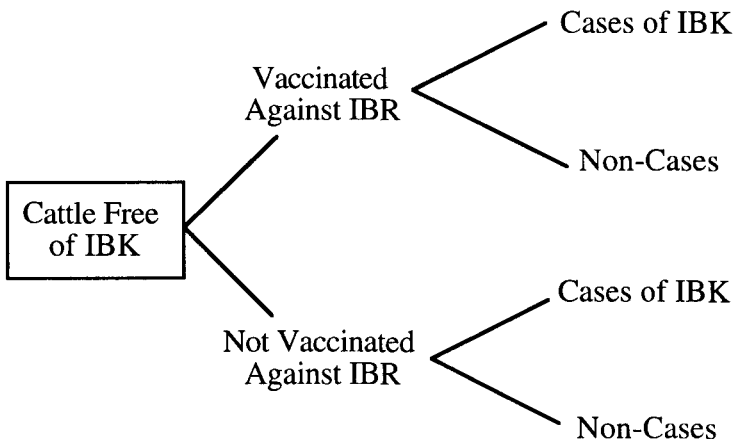


Figure 6.1 A cohort study of the risk of IBR vaccination for IBK.

idence of IBK. These examinations would continue until a sufficient number of cases of IBK had been observed to allow a firm conclusion about the risk of prior vaccination against IBR.

The researchers would be able to measure directly the risk, or incidence, of IBK in unvaccinated and vaccinated cattle and compute a relative risk of IBK by dividing the incidence in vaccinates by the incidence in nonvaccinates. If relative risk exceeded one and was unlikely to have exceeded one by chance ($P < 0.05$), and vaccinated and unvaccinated cattle did not differ substantially with regard to other risk factors for IBK, then we would conclude that vaccination against IBR is a risk factor for IBK. It would still be necessary to decide whether the association is causal.

2. A Case Control Study of IBK Following IBR Vaccination

A case control study of the same question provides a striking contrast to the cohort study described previously. First, the investigator must find a group of cattle suffering from IBK (Figure 6.2). Cases could be identified from the veterinary hospital's medical records or gradually accumulated from cases as they were presented. Since only cases that were serious enough to require medical attention would be included, those that were relatively mild, healed spontaneously or through the producer's care, would not be represented. Since all of the cases would have been seen during ambulatory visits, numerous farms would most likely be represented.

Once the cases are assembled and the diagnosis confirmed, a comparison or control group would be selected. The question that the investigators are asking is whether cattle suffering from IBK are more likely to have been vaccinated against IBR than a similar group of cattle unaffected by the disease. What is meant by similar? Similarity in the cohort study meant membership in the same cohort, e.g., cattle from ranches with similar management practices. A comparable natural cohort for the group of cases receiving ambulatory care is not possible. Therefore, one must be created.

A control group for these cases could be created by matching each case with the first eligible control animal that does not have a prior history of IBK. Control group cattle should also be matched to affected cattle by age, breed, sex, background management procedures and vaccination history (disregarding IBR vaccination). In this way, a group of cattle is assembled that hopefully is similar to cases with respect to factors that might determine risk for IBK, other than prior vaccination against IBR.

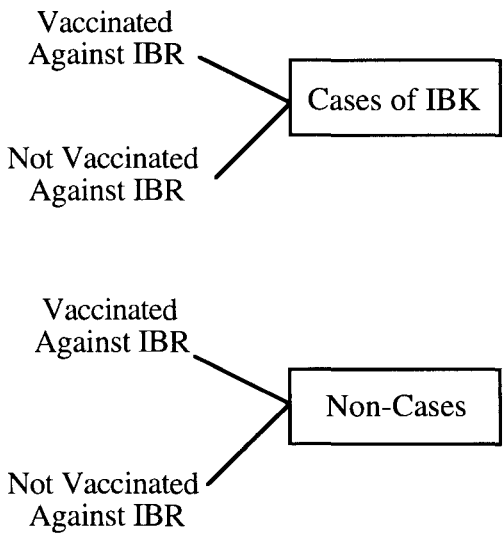


Figure 6.2 A case control study of the risk of IBR vaccination for IBK.

Once the cases and controls have been assembled, the next step is to determine the prevalence of IBR vaccination in cases versus controls. To examine the possible risk of IBR vaccination, each animal's vaccination history must be reconstructed. 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 (Fletcher et al, 1982).

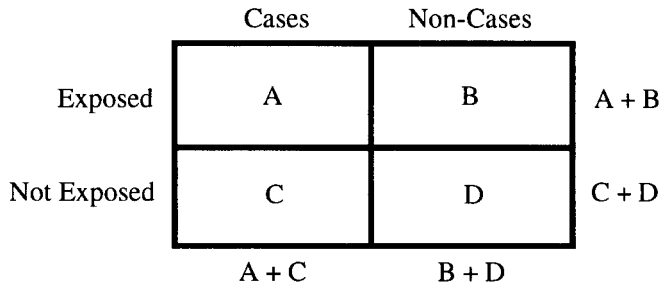
It is the past, not the present, that is important and therein lies a potential for bias in case control studies.

C. 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, incidence of disease. It is, therefore, not possible to calculate relative risk in the usual way. It is possible to obtain an estimate of relative risk in another way, however. 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 6.3). The meaning of the odds ratio is analogous to the relative risk obtained in cohort studies, e.g., the stronger the association between exposure and disease, the higher the odds ratio.

The meaning of the odds ratio is analogous to the relative risk obtained in cohort studies, e.g., the stronger the association between exposure and disease, the higher the odds ratio.

EXAMPLE: Table 6.9 uses data from Table 6.3 on the risk for calves of low serum gamma globulin levels to calculate the odds ratio as if the study were designed as a case control study. Note that the odds ratio is close, but not identical to the relative risk (OR = 7.02, RR = 6.03).



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

Figure 6.3 Two by two table comparing how the strength of the association between exposure and outcome is estimated from cohort versus case control studies.

Table 6.9 Calculation of odds ratio using data from Table 6.3 as if the study were a case control study

	Cases	*Noncases*	*Total*
Exposed (to low gamma globulin levels)	12	61	73
Not exposed (to low gamma globulin levels)	6	214	220
Total	18	275	293

Odds ratio = (12 x 214) ÷ (61 x 6) = 7.02

The relative risk was 6.03.

This is because the odds ratio is only an estimate of relative risk. However, the more infrequent the disease, the more closely the odds ratio approximates the relative risk.

D. 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.

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 assures that both groups have an equal likelihood of exposure to the

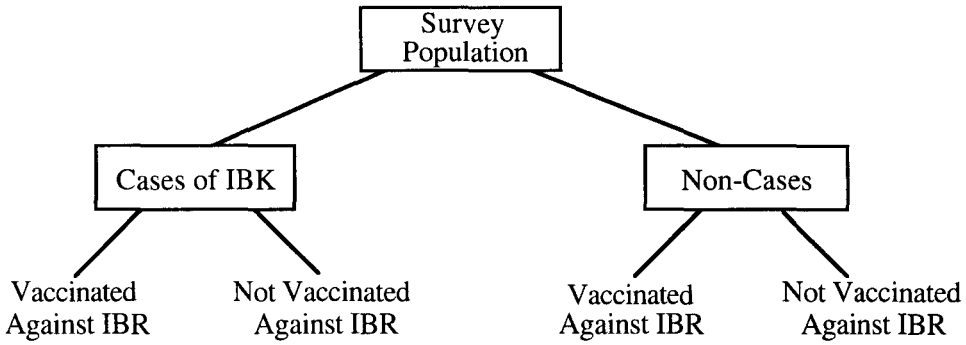


Figure 6.4 A prevalence survey studying the risk of IBR vaccination for IBK.

risk factor of interest. This will facilitate the detection of risk factors which 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.

2. Bias in Measuring Exposure

Measurement bias may occur when the presence of the outcome affects the owner's recollection of the exposure, 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.

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.

VI. 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 measure 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.

Prevalence surveys are especially useful in situations in which we wish to determine which of a number of potential causal 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.

Returning to the issue of the role of IBR vaccination in IBK, a prevalence survey examining this issue would begin and end with a single examination of a large population of beef cattle for IBK and IBR vaccination history (Figure 6.4). The cases would include all cattle found to be suffering from IBK during the survey, and the noncases would include all of the large number of cattle free of the disease. We can be certain that cases and noncases came from the

Table 6.10 Comparison of characteristics of cohort, case control and prevalence survey 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 (i.e., non-cases) not selected – evolve naturally	Controls selected by investigator to resemble cases	Noncases include those free of disease at the 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

From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology - The Essentials*, first edition, Rare Disease. Copyright 1982, The Williams & Wilkins Company. With permission.

same population, but the vaccination history must be reconstructed from interviews and medical records. Additionally, the cases include only those detected, or prevalent, during the examination. Even though the entire population may be sampled, the computed rates are prevalences, not incidences. Since incidence is not being measured, it is preferable to use the odds ratio to estimate risk in prevalence surveys. 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 which does not affect disease incidence but is related to survival of the cases will be associated with disease prevalence in a cross-sectional study (Dicker, 1993). Characteristics of cohort, case control and prevalence survey designs are compared in Table 6.10.

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 each suspected risk factor (age, breed, sex, etc.) among cases is then compared with prevalences for the remaining clinic population 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: Pelvic fractures are an infrequent cause of lameness in horses, with reported frequencies ranging from 0.9% to 4.4% of all lamenesses. Little and Hilbert (1987) reviewed the

Table 6.11 Age, breed and sex distribution of horses with pelvic fractures, compared with the equine clinic population

Risk Factor	*Horses with Pelvic Fractures (% of total)*	*Clinic Population (% of total)*
Age		
0-12 months	7 (37%)	645 (16%)
1-2 years	0	384 (10%)
2-4 years	5 (26%)	660 (17%)
5-7 years	5 (26%)	839 (21%)
> 7 years	2 (11%)	1428 (36%)
Breed		
Saddlebred	6 (31.5%)	130 (3%)
Arabian	3 (15.75%)	1019 (26%)
Quarter Horse	3 (15.75%)	1151 (29%)
Other	7 (37%)	1656 (42%)
Sex		
Male	2 (11%)	826 (21%)
Castrated male	1 (5%)	1058 (27%)
Female	16 (84%)	2043 (51%)
Unknown	0	29 (1%)

Reprinted with permission from Little, C. and Hilbert, B. 1987. Pelvic fractures in horses: 19 cases (1974-1984). *J.A.V.M.A.* 190:1203-1206.

medical records and radiographs of all horses (n = 19) with pelvic fractures seen at a VMTH over a 10-year period and evaluated the contribution of several risk factors. The age, breed and sex distribution of horses with pelvic fractures were *compared with the entire clinic population during the same period* (Table 6.11). The age distribution of the two groups was significantly different, with a greater proportion of younger horses in the fracture group (P = 0.023). The male to female ratio of horses with pelvic fractures was 1 to 5.3 (3:16) compared with the ratio for the entire clinic population of 1 to 1.08 (1884:2043; P = 0.018). The cases included a significantly greater proportion of American Saddlebred horses than did the clinic equine population (P < 0.00001). The higher prevalence of pelvic fractures in young horses (four years old or less) may reflect the more vigorous activities of younger horses. The predominance of females may be associated with differences in the size and form of the pelvis in female versus male horses. The higher prevalence in Saddlebreds could not be explained.

VII. 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.

EXAMPLE: Blood samples were collected from 53 dairy cows with uterine prolapse (cases) and from 53 cows with normal parturition matched by dairy for various management programs (controls). Cows with uterine prolapse had significantly lower ($P < 0.01$) total serum calcium content than did controls, suggesting a cause-effect relationship. Since treatment of prolapse and blood collection were done shortly after the prolapse had occurred, the authors believed that there was little likelihood of hypocalcemia developing after the prolapse and before the time of sampling. Hypothesized mechanisms (biological plausibility) for the association between hypocalcemia and uterine prolapse were (1) prolonged recumbency and tenesmus due to hypocalcemia, thus predisposing to uterine prolapse, (2) reduced uterine tone due to hypocalcemia and (3) delayed involution of the cervix due to hypocalcemia (Risco et al, 1984).

VIII. SUMMARY

An understanding of the concept of risk is fundamental to an understanding of such diverse clinical issues as prognosis, treatment and cause. Factors that are associated with an increased risk of acquiring disease are called risk factors. Exposure to risk factors may occur instantaneously or may be chronic or ongoing.

Risk may be estimated through the use of cohort (prospective), case control (retrospective) or prevalence survey study designs. In a true cohort study a group of individuals that have something in common (the cohort) is assembled and followed over time to detect occurrences of the outcome of interest. True cohort studies can be conducted in two ways. In a concurrent cohort study the cohort is assembled in the present and followed into the future. In a historical cohort study the study group is assembled from past records and followed into their future, usually up to the present.

A survival cohort is the name given to a group of individuals who are assembled at various times in the course of their disease, rather than at the beginning. The name does not imply that survival is being studied, but rather that each individual has survived, or is available for study, after a given period of exposure or disease. 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.

To compare risks in cohort studies, several measures of the association between exposure and disease, called measures of effect, are commonly used. Relative risk, or risk ratio, is the ratio of incidence in exposed individuals to incidence in nonexposed individuals. If no additional risk is associated with exposure to a suspected risk factor, 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 exposure and disease, and is frequently used in studies of disease etiology.

Attributable risk, also known as risk difference, is equal to the incidence of disease in exposed individuals minus the incidence in nonexposed individuals. Attributable risk is the additional incidence of disease among individuals attributable to a risk factor.

Population attributable risk is estimated by multiplying the attributable risk by the prevalence of the 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 which is quite prevalent could contribute more to disease incidence in a population than a stronger risk factor which is rarely present.

The population attributable fraction is estimated by dividing the population attributable risk by the total incidence of 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.

Cohort studies are often impractical due to the relative infrequency of most diseases. Case control, or retrospective, studies look backward to compare the proportion of cases and noncases that were exposed to the factor(s) of interest. 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. The stronger the association between exposure and disease, the higher the odds ratio.

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.

Prevalence surveys share some of the characteristics of cohort and case control studies. Prevalence surveys measure the distribution of risk factors among cases and noncases in a defined population. Since prevalence surveys are cross-sectional and incidence is not being measured, it is preferable to use the odds ratio to estimate risk in prevalence surveys.

Chapter 7

MEASURING AND COMMUNICATING PROGNOSES

I. 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 is either 0% or 100%. Practitioners should avoid statements that can be misconstrued as a contract – a definite statement about an outcome. Clients must be appraised of the probabilities of unfavorable, as well as favorable, outcomes. The object 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).

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 course 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: Metabolic changes associated with diarrhea in neonatal calves include a number of blood biochemical changes. Several investigators have indicated that acidosis and hyperkalemia are major causes of death in many of these diarrheic calves. Kasari and Naylor (1985) evaluated the relative merits of treating acidosis in dehydrated, diarrheic calves using sodium bicarbonate, sodium L-lactate, sodium acetate and saline (sodium chloride) concomitant with parenteral fluid therapy. Thirty-six calves with spontaneously occurring diarrhea and dehydration were randomly assigned to four double-blind experimental fluid groups (nine calves per group) designated "saline control," "lactate," "acetate" and "bicarbonate" groups. Acid-base values and selected hematologic and biochemical values were determined from venous blood samples collected from each calf immediately before and after administration of fluid therapy. Dramatic improvements in base deficit relative to controls were measured in calves receiving lactate, acetate and bicarbonate solutions. The magnitude of the response was also related to

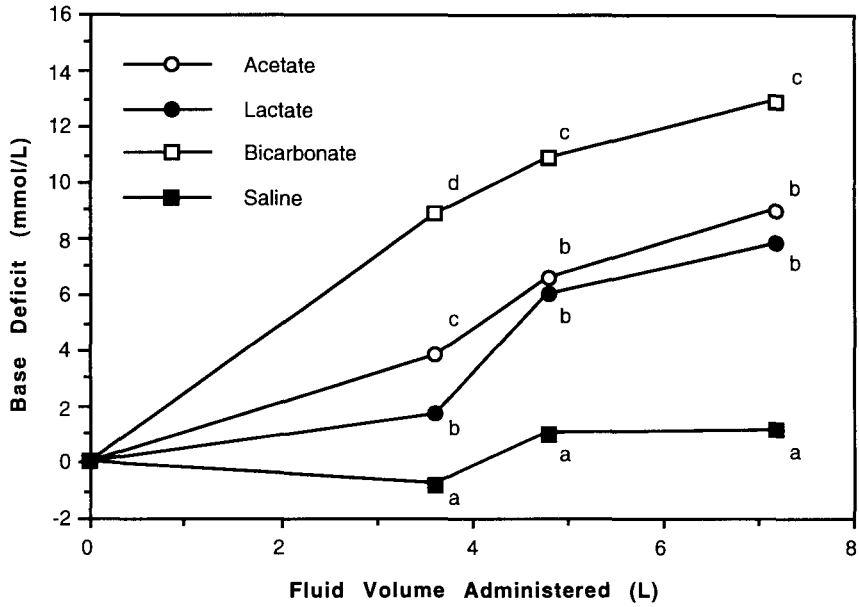


Figure 7.1 Base deficit of 36 dehydrated diarrheic calves (nine calves per group) that received different alkalinizing compounds (50 mmol/L) during extracellular fluid replacement therapy. At a given volume of fluid, means with different letters are significantly different ($P < 0.01$). The initial base deficit was 18.2 ± 1.3 mmol/L. (From Kasari, T.R. and Naylor, J.M. 1985. Clinical evaluation of sodium bicarbonate, sodium L-lactate, and sodium acetate for the treatment of acidosis in diarrheic calves. *J.A.V.M.A.* 187:392-397. With permission.)

volume of fluid administered (Figure 7.1). However, the degree of clinical response of calves to rehydration therapy was directly related to the volume of fluid administered, regardless of the fluid used (Figure 7.2). Despite this, the authors concluded that rehydration of a calf without attention to correcting acidosis via alkalinizing compounds should be avoided.

Wherever possible, prognoses should be assessed in ways that can be perceived by the patient and its owner.

II. 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 medical teaching hospitals and other referral centers may not be representative of cases seen in the typical private practice. Reported cases are often those which had been referred because they were doing badly.

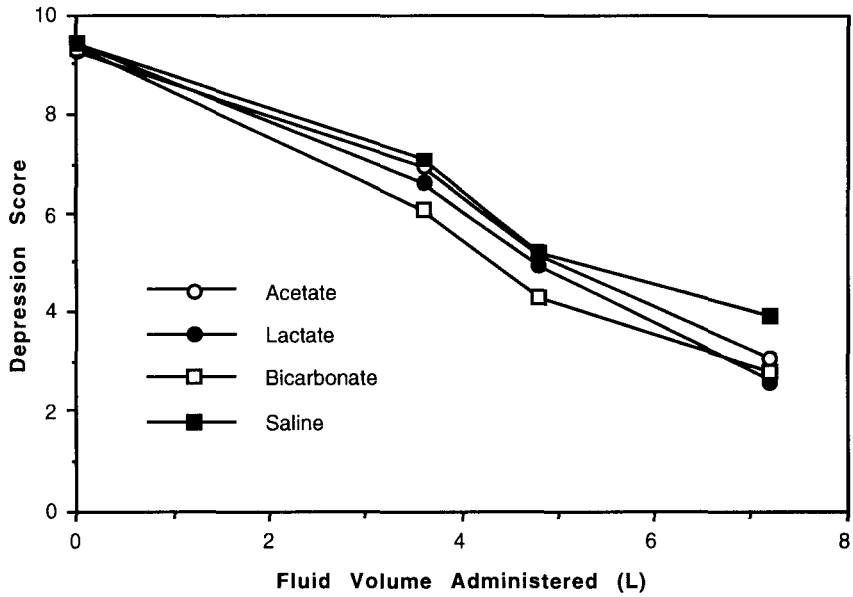


Figure 7.2 Influence of extracellular fluid replacement therapy on depression scores in dehydrated calves. Statistically significant differences were not found between groups, as determined by analysis of variance. (From Kasari, T.R. and Naylor, J.M. 1985. Clinical evaluation of sodium bicarbonate, sodium L-lactate, and sodium acetate for the treatment of acidosis in diarrheic calves. *J.A.V.M.A.* 187:392-397. With permission.)

Reports of prognosis from veterinary medical teaching hospitals and other referral centers may not be representative of cases seen in the typical private practice.

EXAMPLE: A study of the prognosis of feline leukemia virus (FeLV) infection in a cohort of cats with newly acquired infection provided a rare opportunity to study the natural history of the disease (Hardy et al, 1976). Fifty-five clinically normal cats who acquired FeLV infection from household contacts over a 3-month period were followed over time. Over the 2-year follow-up period nine cats were euthanized. Fifty-two percent of the remaining 46 FeLV-infected cats died; 13 (28%) from lymphosarcoma and other FeLV-caused diseases and 11 (24%) from other diseases. Based on data from unmatched controls (McClelland et al, 1980), fewer than 16% of FeLV-free cats would be expected to die over the same time period, and fewer than 1% from lymphosarcoma or FeLV-caused diseases (Table 7.1).

III. PROGNOSIS AS A RATE

It is convenient to summarize the course of disease as a rate. Rates commonly used for this purpose are shown in Table 7.2. 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 a survival cohort of patients. Zero time may be assigned at any point in the course of disease such as onset of signs, diagnosis or treat-

Table 7.1 Mortality among FeLV-infected and -uninfected cats from the time at which infection was acquired

Cause of Death	*Incidence in FeLV-Infected Cats[a] (n = 46)*	*Incidence in Uninfected Cats[b] (n = 512)*
FeLV diseases		
Lymphosarcoma	15.2%	0.6%
Others[c]	13.0%	0.2%
Non-FeLV diseases		
Feline infectious peritonitis	6.5%	1.2%
Others	17.4%	14.1%
Overall	52.2%	16.1%

[a] Based on 2-year follow-up. Source of data: Hardy, W.D., Jr., McClelland, A.J., Zuckerman, E.E., Hess, P.W., Essex, M., Cotter, S.M., MacEwen, E.G., and Hayes, A.A. 1976. Prevention of the contagious spread of feline leukaemia virus and the development of leukaemia in pet cats. *Nature.* 263:326-328.

[b] Based on 3.5-year follow-up. Source of data: McClelland, A.J., Hardy, W.D., Jr., and Zuckerman, E.E. 1980. Prognosis of healthy feline leukemia virus infected cats. In, W.D. Hardy, Jr., M. Essex, and A.J. McClelland (eds), *Feline Leukemia Virus.* Elsevier, New York, pp. 121-126.

[c] Nonregenerative anemias, panleukopenia-like syndrome.

Table 7.2 Rates commonly used to describe a prognosis

Rate	*Definition*
Survival	Percent of patients surviving a defined period of time from some point in the course of their disease
Case fatality	Percent of patients with a disease who die of it
Response	Percent of patients showing some evidence of improvement following an intervention
Remission	Percent of patients entering a phase in which disease is no longer detectable
Recurrence	Percent of patients who experience a return of disease after a disease-free interval

From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology – The Essentials*, first edition, Prognosis. Copyright 1982, The Williams & Wilkins Company. With permission.

Table 7.3 Mortality over 3.5-year follow-up among FeLV-infected and uninfected cats from the time at which infection was diagnosed

Cause of Death	*Incidence in FeLV-Infected Cats (n = 96)*	*Incidence in Uninfected Cats (n = 512)*	*Relative Risk*	*Attributable Risk*
FeLV diseases				
Lymphosarcoma	27.1%	0.6%	45.2	26.5%
Others[a]	7.3%	0.2%	36.5	7.1%
Non-FeLV diseases				
Feline infectious peritonitis	5.2%	1.2%	4.3	4.0%
Others	43.7%	14.1%	3.01	29.6%
Overall	83.3%	16.1%	5.21	67.2%

[a] Nonregenerative anemias, panleukopenia-like syndrome.

Source of data: McClelland, A.J., Hardy, W.D., Jr., and Zuckerman, E.E. 1980. Prognosis of healthy feline leukemia virus infected cats. In, W.D. Hardy, Jr., M. Essex, and A.J. McClelland (eds), *Feline Leukemia Virus.* Elsevier, New York, pp. 121-126.

ment. Consequently the computed rates will depend heavily upon 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.

EXAMPLE: The results of natural disease development over a 3.5-year period in initially healthy, FeLV-infected and uninfected cats is summarized in Table 7.3 (McClelland et al, 1980). The feline cohort in this study differs from that in Table 7.1 in that the duration of infection at the start of this study is not known (i.e., it is a survival cohort). Thus, the interval of follow-up is from time of diagnosis, rather than time of infection. In Table 7.3 the original data have been used to calculate relative and attributable risks. The cause of death has been partitioned into FeLV-related and unrelated diseases. Despite the difference in study design, yearly mortality for FeLV-infected cats in Tables 7.1 (26.1%) and 7.3 (23.8%) is very similar. Yearly mortality for uninfected cats in the same studies was only 4.6%.

Rates, such as those in Table 7.2, 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 we analyze it over time.

IV. SURVIVAL ANALYSIS

When interpreting a prognosis, we would like to know the likelihood, on the average, that patients with a given condition will experience an outcome at any point in time. When prognosis is expressed as a summary rate it does not contain this information. However, a method called *survival analysis* provides information about average time to event for any point in the course of disease. The plotted data are referred to as a *survivorship curve.*

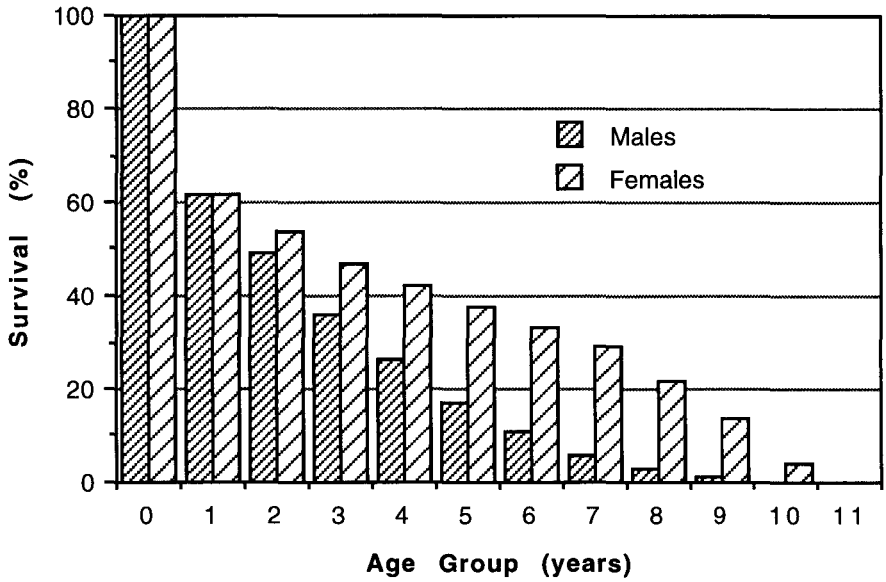


Figure 7.3 Survivorship of White-Tailed Deer. Source of data: Spain, J.D. 1982. *BASIC Microcomputer Models in Biology.* Addison-Wesley, p. 114.

Similar overall rates may cover up important differences in prognosis over the course of a disease.

A. SURVIVAL OF A COHORT

The most direct way of learning about survival is to assemble a cohort of patients with the condition of interest and periodically count the number remaining throughout the course of their illness. Life expectancy, the expected survival of presumably "normal" individuals, is a form of prognosis. Indeed, the term "terminal" is not unique to diseases – life itself follows a terminal course, which begins at birth. Knowledge of the expected survival of normal individuals provides a baseline for comparison with their diseased counterparts.

1. Steady-State Population Models

When populations are in a steady state, i.e., constant rates of birth and death with no migration in or out of the population over the life span of the individuals, then the age frequency distribution can be used to estimate the survival of a cohort of the population. This is depicted graphically in Figure 7.3 where survivorship curves for white-tailed deer have been derived from a population model of a Michigan herd (Spain, 1982). The additional insight provided by survivorship curves is apparent when we compare the survival of male versus female deer. The survival rates for male and female deer are identical through 1 year of age, but they diverge markedly thereafter. The reduced survival in the male population over 1 year of age is due primarily to hunting pressure.

2. Vital Statistics Data

Many populations are not in a steady state. For example, we are all familiar with the ups and downs of the birth rate in the U.S. population and have heard many accounts of the effect of the "baby boomers" on the demand for teachers, goods and services and the housing market. Changes in the birth and death rate over time are reflected in statistics on the age frequency distribution of the U.S. population. However, the death rate for any particular year can be used

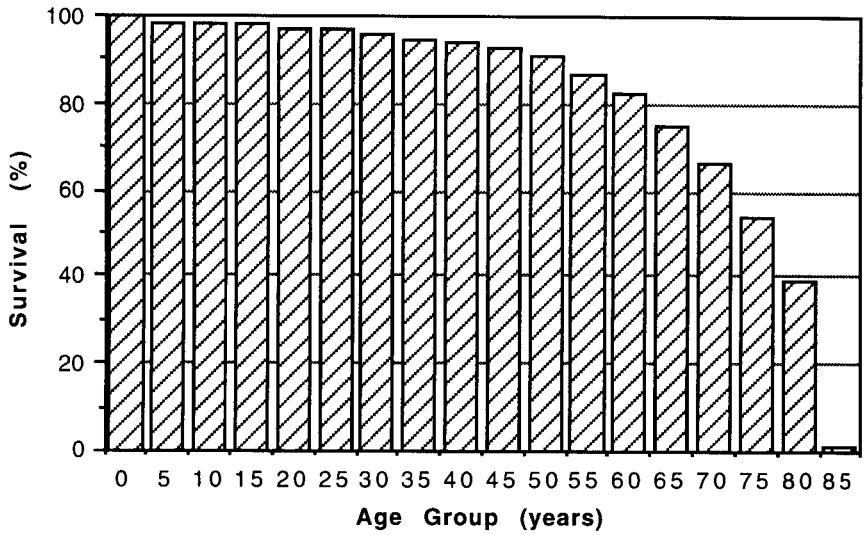


Figure 7.4 Survivorship curve for the U.S. population for 1976. (Source of data: Bureau of Census. 1978. Expectation of life and mortality rates, by race, age, and sex: 1976. *Statistical Abstract of the United States.* U.S. Department of Commerce.)

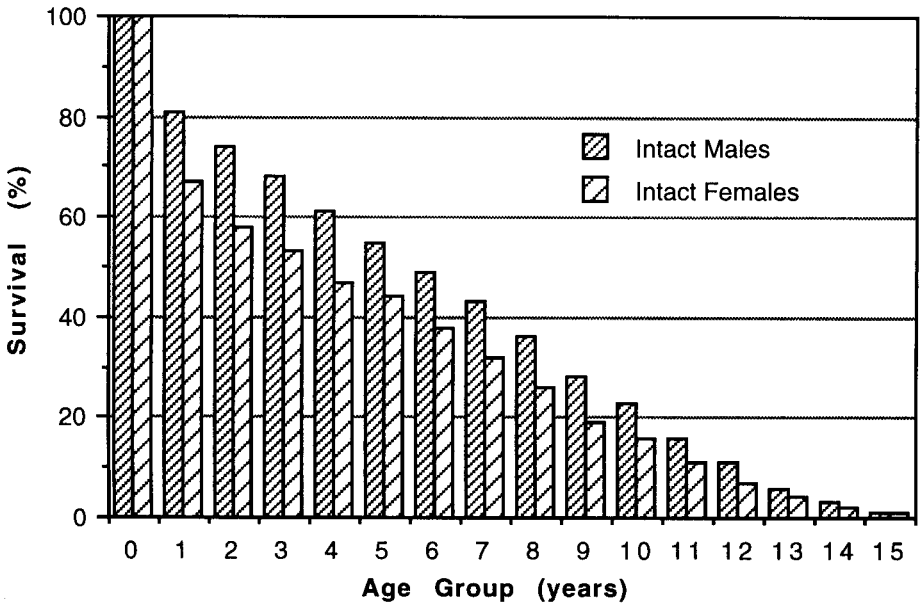


Figure 7.5 Canine survivorship curve for intact males and females. (Source of data: Bronson, R.T. 1982. Variation in age at death of dogs of different sexes and breeds. *Am. J. Vet. Res.* 43:2057-2059.)

to estimate a survivorship curve for the human population. Since a rate is used, rather than absolute numbers of deaths (dangling numerators), the resulting survivorship data are unaffected by the number of individuals in each age class. Figure 7.4 depicts a human survivorship curve based on the age class death rates of the 1976 U.S. population (Bureau of Census, 1978).

Table 7.4 Analysis of data from a cohort of cats undergoing chemotherapy for advanced mammary adenocarcinoma where all were observed until death (complete follow-up)

Original Data		*Survival of the Cohort*			
	Survival Time	*Time Interval*		*Remaining*	
Patient	*(d)*	*(d inclusive)*	*Deaths*	*No.*	*Percent*
4	4	0	0	11	100
5	45	30	1	10	91
6	47	60	2	8	73
7	67	90	1	7	64
8	106	120	1	6	55
9	149	150	1	5	45
10	170	180	2	3	27
11	180	210	1	2	18
12	182	240	0	2	18
13	283	270	0	2	18
14	344	300	1	1	9
		330	0	1	9
		360	1	0	0
			11		

Survival times from Table 6.5.
Mean = 143; median = 149.

Source of data: Jeglum, K.A., de Guzman, E., and Young, K.M. 1985. Chemotherapy of advanced mammary adenocarcinoma in 14 cats. *J.A.V.M.A.* 187:157-160.

Unfortunately, comparable vital statistics data are not routinely collected for animal populations. The closest that we can come is the distribution of age at death. The following example was taken from diagnostic laboratory data that were used to estimate the longevity of different breeds of dogs (Bronson, 1982). There are a number of biases inherent in these data. The survival analysis that follows hinges on two assumptions: (1) the age distribution of dogs presented for necropsy is representative of all dogs dying in the area and (2) the population is in a steady state. Figure 7.5 is based on the assumption that a dog that died in a given age interval would have been alive during all preceding intervals (Lebeau, 1953), an assumption inherent to the Reed and Muench method of estimating the 50% lethal dose. Despite the likely effect of bias on the resulting survivorship curve, it is apparent that the pattern of canine survival is markedly different from that of human beings. This should emphasize the inaccuracy of estimating the canine-human age equivalence solely on maximum life span. Dividing 85 by 15 = 5.7 years, suggesting that one year of a dog's life is equal to 5.7 years in the life of a human being. Actually, after one year only about 70% of dogs are alive versus 98% of humans at 5.7 human years.

3. Clinical Trials

Clinical trials frequently describe the prognosis for diseased patients with or without treat-

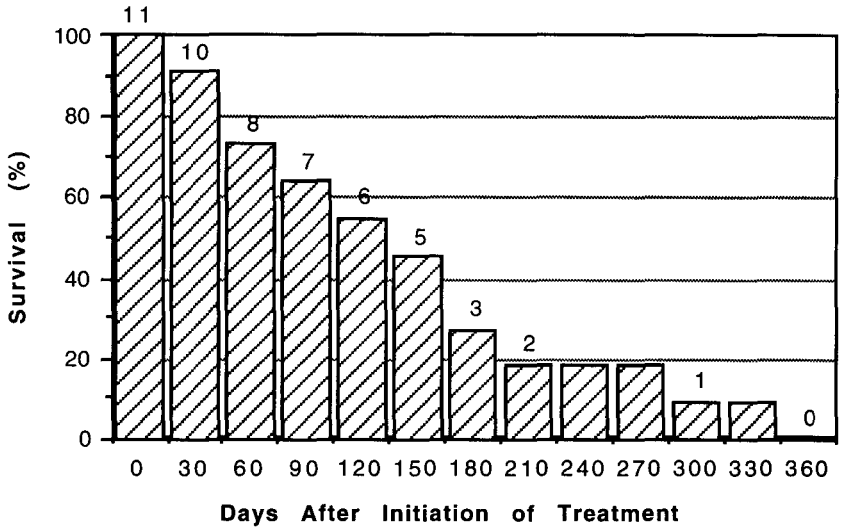


Figure 7.6 Survivorship curve for a cohort of 11 cats following chemotherapy for advanced mammary adenocarcinoma. Numbers above bars correspond to the number of cats remaining in the cohort. (Source of data: Jeglum, K.A., de Guzman, E., and Young, K.M. 1985. Chemotherapy of advanced mammary adenocarcinoma in 14 cats. *J.A.V.M.A.* 187:157-160.)

ment. The results may be expressed as rates, as mentioned previously, or depicted as survivorship curves. Frequently sufficient information is available for construction of survivorship curves, but it is "hidden away" in the text of the report.

EXAMPLE: In Chapter 6, data were presented from a survival cohort of cats with advanced mammary adenocarcinoma in which the chemotherapeutic cycle was repeated every 21 days until death (see Table 6.5). If we exclude the three cats for which no follow-up data were available, we are left with a cohort of 11 cats from which a survivorship curve can be constructed. It is important to note that all 11 cats were followed until the outcome (death) occurred. The original data are analyzed in Table 7.4 along with the resulting survivorship curve (Figure 7.6).

Note that the results can be expressed over fixed time intervals (as in this case) or time to event (death). The former was chosen to simulate the results of a monthly checkup of patients; however, the latter would actually have provided a more accurate representation of the data. The number of individuals on which values for each interval are based is indicated above the interval. The median survival was 149 days, which means that half of the patients survived for this period of time. The mean value of 143 days implies that the average patient would survive this period of time. The median is a better expression of prognosis since the mean value is influenced by extreme values.

B. LIFE TABLE ANALYSIS

Maintaining the integrity of a cohort is often difficult in clinical practice because (1) patients ordinarily become available for a study over a period of time, thus resulting in variable time of follow-up, and (2) patients may drop out of the study before the end of the follow-up period. *Life table analysis* can be used to more efficiently use follow-up data, regardless of the time at which an individual enters or leaves a study. Life table analysis, also known as the *actuarial method*, has been used extensively by the insurance industry.

Table 7.5 Original data from follow-up study of cats treated surgically for hemangiosarcoma

Group	*Time to Event (weeks)*
Still alive at last follow-up	18, 19, 40, 77, 90, 112
Died during follow-up	6, 13, 15, 20, 27, 32, 35, 75, 86

From Scavelli, T.D., Patnaik, A.K., Mehlhaff, C.J., and Hayes, A.A. 1985. Hemangiosarcoma in the cat: retrospective evaluation of 31 surgical cases. *J.A.V.M.A.* 187:817-819. With permission.

Table 7.6 Life table with time-to-event intervals using data from Table 7.5 on feline hemangiosarcoma

	No. of Events			*Survival*	
Interval (weeks)	*Censored*	*Death*	*At Risk*	*Interval (%)*	*Overall (%)*
0	0	0	15	–	100
6	0	1	15	93	93
13	0	1	14	93	87
15	0	1	13	92	80
20	2	1	10	90	72
27	0	1	9	89	64
32	0	1	8	88	56
35	0	1	7	86	48
75	1	1	5	80	38
86	1	1	3	67	26
90	0	0	2	100	26
112	1	0	1	100	26

Source of data: Scavelli, T.D., Patnaik, A.K., Mehlhaff, C.J., and Hayes, A.A. 1985. Hemangiosarcoma in the cat: retrospective evaluation of 31 surgical cases. *J.A.V.M.A.* 187:817-819.

With the life table method, the probability of surviving over each time interval is calculated by dividing the number of patients surviving by the number at risk of dying during the 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 calculation for that interval. If there have been no deaths over an interval, then the probability of surviving remains the same and is not recalculated. The chance of surviving to any point in 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.

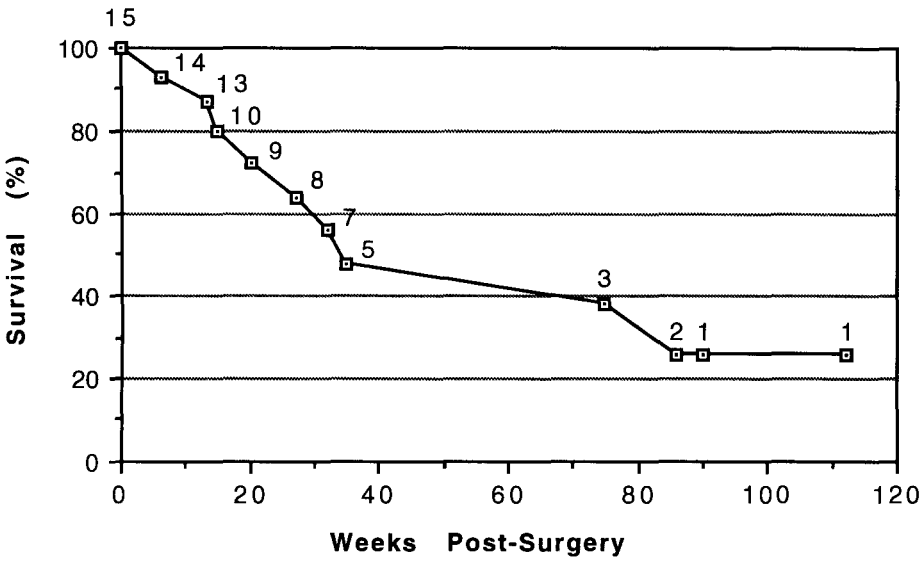


Figure 7.7 Survivorship curve depicting postoperative survival of 15 cats being treated for hemangiosarcoma. Numbers above data points correspond to number of cats remaining in the cohort. (Source of data: Table 7.6.)

The major difference between analysis of cohort data with complete follow-up data, as depicted in Table 7.4, and life table analysis is that in the latter the number of individuals at risk over each interval must be adjusted for individuals who drop out of the study.

EXAMPLE: Hemangiosarcoma, also known as hemangioendothelioma and angiosarcoma, is a malignant neoplasm originating in the endothelium of blood vessels. It develops commonly in the dog, but reports of hemangiosarcoma in the cat are rare. During retrospective analysis of medical records in a veterinary hospital, 31 cases of feline hemangiosarcoma were identified in which therapeutic surgery was performed (Scavelli et al, 1985). Owners were contacted for follow-up information from which postsurgical survival time data were obtained for 20 of the 31 cats. Of these, three were euthanized at surgery and two in the first postoperative week. Nine of the remaining 15 cats died over the 112-week postoperative follow-up period, while six cats were still alive from 18 to 112 weeks post-surgery. The original data appear in Table 7.5.

Survival analysis of these data is complicated by *censored observations*, e.g., patients having incomplete follow-up (Thomas et al, 1977). In order to accommodate censored observations, it is necessary to restructure the data into the life table form depicted in Table 7.6. The difference between this study and the cohort analysis described in Table 7.4 and Figure 7.6 is that several cats were not followed over the duration of the study because they were added sometime between its initiation and the end. Consequently, the population at risk when each event (death) occurred was adjusted for previous deaths and loss to follow-up. Thus, even though all cats were not followed for the same period of time, each contributed to the analysis for the period that it remained in the study. The resulting survivorship curve appears in Figure 7.7.

The life table approach can be used to describe other outcomes of disease besides death, e.g., recurrence of tumor, remission duration, rejection of graft or reinfection, and to identify prognostic factors for these outcomes. In fact, the frequency of any event can be studied by means of life tables, as long as the event is dichotomous (i.e., either/or), and the event can occur only once during the follow-up period. The following two examples illustrate the use of

Table 7.7 Duration of observation versus outcome for horses undergoing corrective shoeing for navicular disease

Months	*Number Observed*	*Number Not Lame*	*Number Lame*
12-18	3	3	0
18-24	2	1	1
24-30	1	0	1
30-36	2	2	0
36-42	7	6	1
42-48	8	7	1
48-54	13	12	1

From Turner, T.A. 1986. Shoeing principles for the management of navicular disease in horses. *J.A.V.M.A.* 189:298-301. With permission.

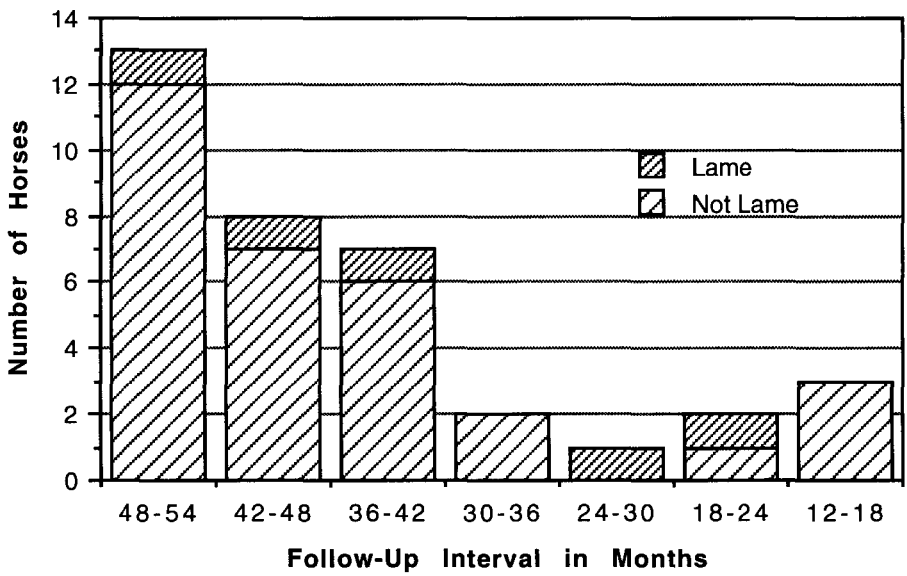


Figure 7.8 Graphic presentation of data from Table 7.7 describing the response of horses to corrective shoeing for navicular disease. (Source of data: Table 7.7.)

life table analysis to evaluate remission duration (following corrective shoeing for navicular disease in horses) and to identify prognostic factors (for survival in dogs afflicted with multiple myeloma).

EXAMPLE: Navicular disease is a commonly diagnosed cause of lameness in horses and has been reported to cause one third of all chronic forelimb lamenesses. Navicular disease was diagnosed between August 1979 and November 1982 in 36 horses (Turner, 1986). Each was treated by corrective shoeing. Shoes were reset every four to six weeks. Treatment was con

Table 7.8 Life table with time-to-event intervals using data from Table 7.7 on navicular disease in the horse

Interval (mo)	*No. of Events*		*At Risk*	*In Remission*	
	Censored	*Lame*		*Interval (%)*	*Overall (%)*
12-18	---	0	36	100	100
18-24	3	1	33	97	97
24-30	1	1	31	97	94
30-36	0	0	30	100	94
36-42	2	1	28	96	90
42-48	6	1	21	95	86
48-54	7	1	13	92	79

Source of data: Turner, T.A. 1986. Shoeing principles for the management of navicular disease in horses. *J.A.V.M.A.* 189:298-301.

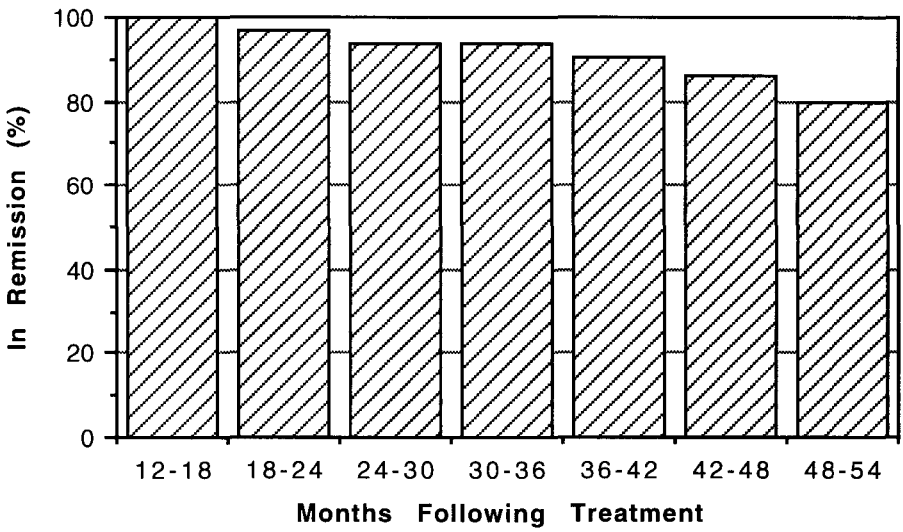


Figure 7.9 Graphic presentation of data from Table 7.8 in which data describing the response of horses to corrective shoeing for navicular disease have been submitted to life table analysis. (Source of data: Table 7.8.)

sidered successful if lameness could not be detected at the trot at hand and the horse was competing at or above its prelameness level. Thirty-one horses were free of lameness as of February 1984 when the study was concluded. Follow-up periods thus ranged from 12 to 54 months. The original data are presented in Table 7.7 and summarized in Figure 7.8, where the number and disease status of horses for each follow-up interval is presented.

Horses with longer follow-up periods are depicted first to represent their relative time of entry into the study. The data have been reworked in Table 7.8 to facilitate construction of the survivorship curve in Figure 7.9, which depicts the duration of the disease-free condition

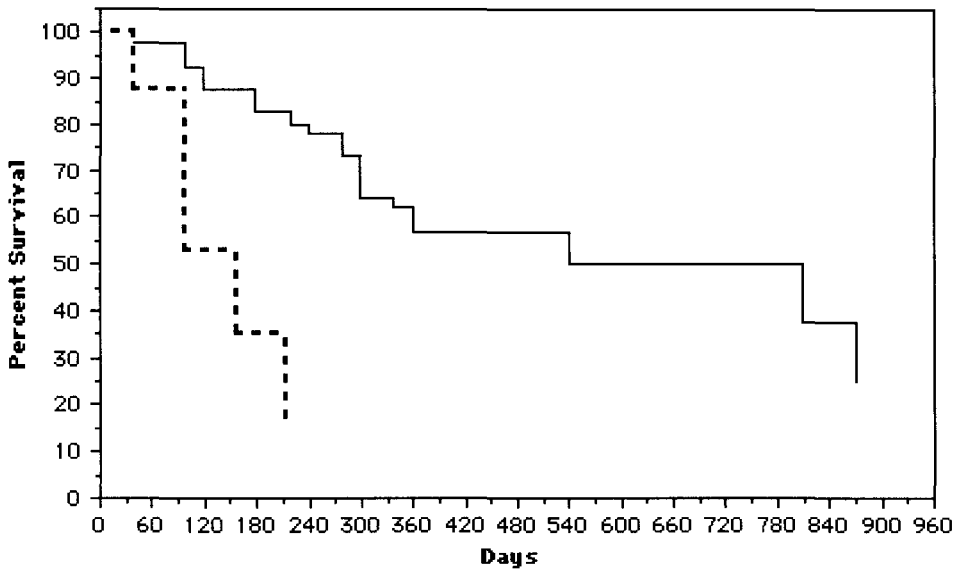


Figure 7.10 Prognostic factors for multiple myeloma in the dog. Survival time of treated dogs based on calcium concentration. Ca < 11.5g/dl (n = 31), —— ; Ca > 11.5 g/dl (n = 6), - - - (P = 0.002). (From Matus, R.E., Leifer, C.E., MacEwen, E.G., and Hurvitz, A.I. 1986. Prognostic factors for multiple myeloma in the dog. *J.A.V.M.A.* 188:1288-1292. With permission.)

(remission duration) following corrective shoeing. The success of shoeing was dependent on the duration of lameness before treatment. Evaluation of clinical trials with this and other types of bias is discussed in Chapter 8.

EXAMPLE: Forty-nine dogs with multiple myeloma were monitored for at least 30 days after diagnosis to establish prognostic criteria (e.g., identify prognostic factors) for the disease based on biological behavior of the tumor (Matus et al, 1986). Of these, 37 (group 1) were treated with alkylating agents combined with prednisone. Twelve dogs (group 2) were given only prednisone as palliative treatment. Assignment to treatment groups was made on the basis of owner compliance and not on clinical stage of disease or performance status of the dog (a potential source of bias). Additional supportive treatment was administered as necessary. Specific therapy prolonged survival (P = 0.04). Hypercalcemia (P = 0.02; Figure 7.10) and Ig light chain proteinuria (Bence Jones proteins; P = 0.04) were significantly associated with shorter median survival times in treated dogs. Sex, monoclonal Ig class, increased serum viscosity and azotemia did not correlate significantly with prognosis (P > 0.05).

C. 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 if Figures 7.6 and 7.7 are compared. The former is based on follow-up of a cohort of individuals with no censored observations. Consequently, the number of individuals remaining at any point on the curve can be estimated by multiplying the percent survival at this point by the number of individuals initially present. In contrast, if we multiply 26% survival by the 15 cats initially present in Figure 7.7, we obtain four cats. Actually, the 26% survival figure

Table 7.9 Numeric equivalents for 16 literal prognoses based on the response of 47 large and small animal practitioners

Prognostic Term or Phrase	*No. of Responses*	*Numeric Designation of Probability of Recovery**		
		Mean	*± S.D.*	*Range*
Terminal	45	0.11	0.38	0-2
Incurable	43	0.21	0.51	0-2
Horrible	41	0.80	0.84	0-3
Grave	47	0.96	0.86	0-3
Dismal	41	1.22	0.88	0-3
Very poor	46	1.96	0.99	0-5
Poor	47	2.64	1.01	0-5
Unfavorable	46	2.78	1.47	0-5
Guarded	46	3.83	1.73	1-8
Not so good	42	3.93	1.54	2-9
Fair	47	5.79	1.59	2-10
Not too bad	42	7.10	1.51	6-10
Favorable	46	8.07	0.83	6-10
Good	47	8.32	0.78	6-10
Very good	46	8.96	0.70	7-10
Excellent	47	9.83	0.38	9-10

*Recovery was defined as absence of disease-related signs for at least one year after appropriate treatment/management.

From Crow, S.E. 1985. Usefulness of prognoses: qualitative terms vs quantitative designations. *J.A.V.M.A.* 187:700-703. With permission.

is based on only one cat, as the others were not available for the entire 112-week follow-up period.

Second, the number of individuals at risk declines as we move from left to right along the survival curve. Consequently, our estimates of the probability of survival depend on what happens to fewer and fewer individuals. A single event towards 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 appears 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.

V. 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.

Table 7.10 Numeric designation for probability of recovery from 22 common illnesses of small animals

Prognostic Term or Phrase	*No. of Responses*	*Numeric Designation of Probability of Recovery**		
		Mean	*± SD*	*Range*
Fleabite dermatitis	20	7.80	2.89	0-10
Otitis externa	20	7.40	3.12	1-10
Hypoadrenocorticism	20	7.25	2.43	2-10
Epilepsy	20	6.30	2.96	0-10
Intervertebral disk disease	19	6.22	2.94	0-9
Diabetes mellitus	19	5.79	2.90	1-9
Hyperadrenocorticism	19	5.68	2.96	2-8
Atopic dermatitis	19	5.21	3.44	0-10
Exocrine pancreatic insufficiency	20	5.20	3.40	0-10
Chronic bronchitis	18	5.06	3.15	0-10
Collapsing trachea	19	4.89	3.13	0-9
Mammary carcinoma	19	4.63	2.77	1-10
Glaucoma	19	4.53	3.13	0-10
Mitral insufficiency with congestive failure	19	4.21	2.57	0-8
Granulomatous colitis	19	3.89	2.54	0-8
Chronic active hepatitis	19	3.00	2.29	0-7
Nasal aspergillosis	18	3.00	2.54	
Distemper	20	2.85	2.78	0-9
Lymphosarcoma	20	2.75	2.36	0-8
Cardiomyopathy	18	2.33	1.91	0-7
Chronic progressive renal disease	18	2.05	1.67	0-6 0-5
Osteosarcoma	21	1.52	1.63	0-6

*Recovery was defined as absence of disease-related signs for at least 1 year after appropriate treatment/management.

From Crow, S.E. 1985. Usefulness of prognoses: qualitative terms vs quantitative designations. *J.A.V.M.A.* 187:700-703. With permission.

EXAMPLE: Table 7.9 summarizes the responses of 47 large and small animal practitioners at a university teaching hospital who were asked to designate numeric equivalents for each of 16 literal terms, on a scale of 0 to 10. The number 0 was assigned to no probability of recovery and each increment of 1 represented a 10% probability of recovery. Recovery was defined as absence of disease-related signs for at least one year after appropriate treatment/management (Crow, 1985). Small animal practitioners were also asked to apply the same numeric scale to 22 common illnesses of dogs and cats, for the purpose of evaluating the disorders with respect to an animal's chances for recovery. The results are summarized in Table 7.10.

Because of the considerable overlap of terms in Table 7.9, the author suggested that veterinarians use the prognostic terms listed in Table 7.11 to express prognoses.

Table 7.11 Qualitative terms for clinical outcomes

Prognosis	*Probability of Recovery (%)*
Excellent	90-100
Good	70-89
Fair	40-69
Poor	10-39
Grave	0-9

From Crow, S.E. 1985. Usefulness of prognoses: qualitative terms vs quantitative designations. *J.A.V.M.A.* 187:700-703. With permission.

VI. SUMMARY

Prognosis is a prediction of the expected outcome of disease with or without treatment. A prognosis should include (1) variability in course 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.

The natural history of a disease describes its evolution without medical intervention. 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. Reports of prognosis from veterinary medical teaching hospitals and other referral centers may not be representative of cases seen in the typical private practice. Reported cases are often those which had been referred because they were doing badly.

It is convenient to summarize the course of disease as a rate. All rates used for this purpose 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.

Survival analysis can be used to obtain information about the average time to event for any time in the course of disease. The plotted data are referred to as a survivorship curve. The most direct way of learning about survival is to assemble a cohort of patients with the condition of interest and periodically count the number remaining throughout the course of their illness.

Maintaining the integrity of a cohort is often difficult in clinical practice because (1) patients frequently drop out of the study before the end of the follow-up period, and (2) patients ordinarily become available for a study over a period of time, thus prolonging the duration of the study. Data on patients with incomplete follow-up are referred to as censored observations. Life table analysis can be used to more efficiently use follow-up data, regardless of the time at which an individual enters or leaves a study. With the life table method, 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 over the corresponding time interval by the probability of surviving up to the beginning of that interval.

The life table approach can be used to describe other outcomes of disease besides death,

e.g., recurrence of tumor, remission duration, rejection of graft or reinfection, and to identify prognostic factors for these outcomes.

Several points must be considered when interpreting survival curves. First, since the data includes censored observations, the percentage of individuals at each data point may not be equivalent to the actual number of individuals remaining in the study. Second, the number of individuals at risk declines as we move from left to right along the survival curve. As a result, we can have less confidence in our estimates of survival toward the end of the survival curve. Finally, the tailing of survival curves may be due to fixed rates of survival being applied to a diminishing number of individuals.

The use of qualitative terms to express chances of success or failure is inherently ambiguous. Furthermore, veterinarians frequently do not agree on the prognosis for many common illnesses. Unfortunately for veterinary clinicians, there is no definitive source of prognostic information about diseases of domestic animals. There is a clear need for studies of prognosis in veterinary medicine.

Chapter 8

DESIGN AND EVALUATION OF CLINICAL TRIALS

I. 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 authors' 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 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" (Anonymous, 1980).

II. 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 individuals (or their owners) that adhere to the prescribed treatment regimen. Thus an efficacious treatment could be ineffective due to poor compliance.

III. CLINICAL TRIALS: STRUCTURE AND EVALUATION

Practitioners initiate an observational study of treatment 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 regimens is usually required. The clinical trial is a cohort study 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. Patients 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 which emerge between the two groups over time are attributed to the treatment. Virtually any parameter can be

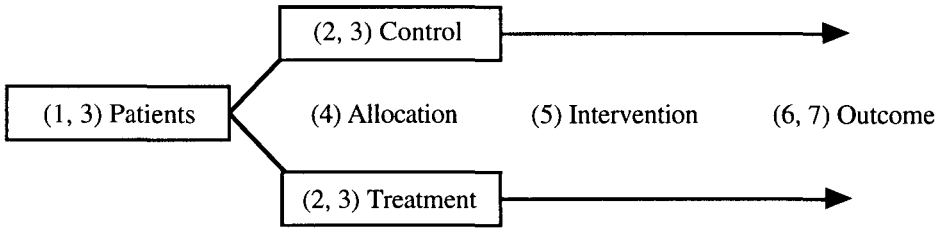


Figure 8.1 Design and potential sources of bias (Table 8.1) in clinical trials. (From Fletcher, R.H., Fletcher, S.W., and Wagner, E.H., *Clinical Epidemiology – The Essentials*, first edition, Treatment. Copyright 1982, The Williams & Wilkins Company. With permission.)

used to measure and express the outcome of a clinical trial. In veterinary medicine the outcome is often expressed in terms of productivity or economic benefit, rather than the health status of individuals.

The clinical trial is a cohort study specifically designed to facilitate the measurement of treatment effects, free of extraneous variables.

Many factors can affect the outcome of cohort studies of risk, prognosis and treatment. These generally originate from one of three sources:

(1) *Assembly bias.* Assembly bias occurs when the criteria for inclusion of patients in a study do not assure uniformity of individuals.
(2) *Migration bias.* Migration bias occurs when patients that leave a study (censored observations) are systematically different from those that remain.
(3) *Measurement bias.* Measurement bias occurs when uniform standards for measurement of clinical events cannot be maintained over time.

The criteria outlined in Table 8.1 have proved 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.

Many factors can affect the outcome of cohort studies of risk, prognosis and treatment. These generally originate from assembly bias, migration bias, or measurement bias.

A. 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.

B. UNCONTROLLED CLINICAL TRIALS

In uncontrolled clinical trials the effects of treatment are assessed by comparing patients'

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 treated 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?

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 relatively unimportant. 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 which 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: Canine ehrlichiosis is a tick-borne rickettsial disease of dogs characterized by fever, pancytopenia, particularly thrombocytopenia, hemorrhage and persistent infection (Smith, 1977). During the initial, acute phase of the disease, clinical signs (nasolacrimal discharge, crusting of the nares, leukopenia) resemble those of several other infectious diseases, particularly canine distemper. Routine hemograms are consistent with this diagnosis. Consequently, veterinarians are seldom prompted to prepare Giemsa-stained buffy coat smears and look for the occasional *Ehrlichia*-infected monocyte, which is pathognomonic for the disease. The natural history of the disease is such that most dogs undergo an uneventful recovery from the acute phase of the disease, regardless of treatment. Consequently, uncontrolled clinical trials of any therapeutic regimen for the disease, correctly diagnosed or not, are likely to be

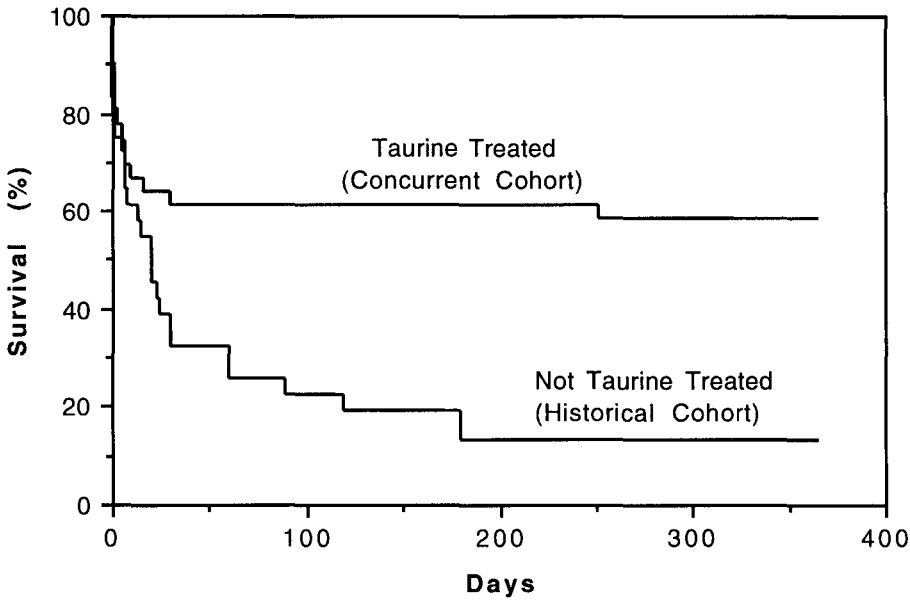


Figure 8.2 Effect of taurine supplementation upon survival of cats from the time of diagnosis of dilated cardiomyopathy. Fifty-eight percent of 36 taurine treated cats (concurrent cohort) survived a year or more versus only 14% of 31 untreated cats (historical cohort). (Source of data: Pion, P.D., Kittleson, M.D., Thomas, W.P., Delellis, L.A., and Rogers, Q.R. 1992. Response of cats with dilated cardiomyopathy to taurine supplementation. *J.A.V.M.A.* 201:275-284. With permission.)

favorable if initiated during the acute phase of infection. More severe complications usually develop months later, during the chronic phase of canine ehrlichiosis.

C. 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 may be the only alternative when it is ethically inappropriate to withhold a promising new treatment from client-owned animals.

EXAMPLE: Dilated cardiomyopathy (DCM) in cats has always been considered a progressive, irreversible condition with a grave prognosis, despite medical intervention. Pion et al (1992) observed rapid reversal of signs following taurine supplementation of affected cats, and designed a clinical trial to evaluate the long-term benefits of administering taurine to cats with DCM. A concurrent cohort of 37 taurine-treated DCM cats (treatment group) was compared with a historical cohort of 33 DCM cats (control group) who had been treated with conventional therapy, before the role of taurine was suspected. The latter group was assembled from medical records by identifying cats with an echocardiographically confirmed diagnosis of DCM. Treatment and survival time data were obtained from the medical records, and verified and supplemented through follow-up telephone interviews with clients. According to treatment records most control cats had received digoxin and furosemide. Cats in the treatment

group with evidence of congestive heart failure were treated symptomatically with a combination of digoxin, diuretics, angiotensin-converting-enzyme inhibitors, and pleurocentesis. All treatment group cats received oral taurine supplementation initially. Medications other than taurine were discontinued in the treatment group as clinical improvement became evident. Taurine supplementation was discontinued once echocardiographic improvement occurred and plasma levels were maintained through feeding of commercial cat food containing additional taurine.

The survival curves for the two groups (Figure 8.2) diverged markedly within a few weeks after the initiation of taurine supplementation of treatment group cats. Twenty-one (58%) of 36 taurine treated cats with a known outcome survived for at least one year versus only 4 (14%) of 31 untreated cats with a known outcome. Although the differences in the survival curves of the groups were statistically significant, differences in the nature of supportive medications given the two groups confounded the interpretation of the results. Based on historical data the authors discarded the possibility that medications other than taurine were responsible for the improved survival. They also pointed out that it would have been "ethically inappropriate" to withhold taurine supplementation from a concurrent control group of client-owned animals once the beneficial effects of taurine became apparent.

D. ALLOCATING TREATMENT

When concurrent controls are used, assignment to treatment or comparison groups can be done in several ways.

(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, patient 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 by means of 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 groups. If the number of patients is small, the investigator can compare the distribution of a number of patient characteristics among the treatment 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.

E. 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. 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.

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 treatment guidelines.

F. 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, most clinical outcomes are subject to the interpretations 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 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.

G. STATISTICAL ANALYSIS

Many reports of clinical trials end by concluding that a treatment offered a "significant" improvement over existing techniques or controls. Any time 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 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.

IV. CASE STUDIES

The following five case studies are representative of articles on treatment appearing in veterinary practice journals. All present clinically useful information, but potential biases should be taken into account before the information is applied in practice. The evaluation of each article, according to the criteria outlined in Table 8.1, is summarized in Table 8.2. The appropriateness of statistical analyses used in each study is discussed in the following chapter.

A. CASE 1: TREATMENT OF EQUINE COLIC (GINGERICH ET AL, 1985)

1. Background

Effective analgesia is paramount in horses experiencing acute abdominal pain ("colic") to prevent self-inflicted trauma and intestinal displacement and to serve as an aid in performing diagnostic procedures. Among the most common causes of colic are intestinal impaction, intestinal hypermotility, flatulence, postpartum pain, torsion, hypomotility and ulcers.

2. Study Design

Thirteen equine practitioners from various localities in the United States participated in a clinical trial of a new analgesic, butorphanol tartrate (Torbugesic, Bristol), to relieve the pain of equine colic. Subjects (n = 206) were selected on the basis of clinical signs of colic, which were categorized as severe (35%), moderately severe (46%) or mild (19%). Prognosis for recovery was good in 65% of cases, fair in 17% and poor in 18%. The duration of colic before

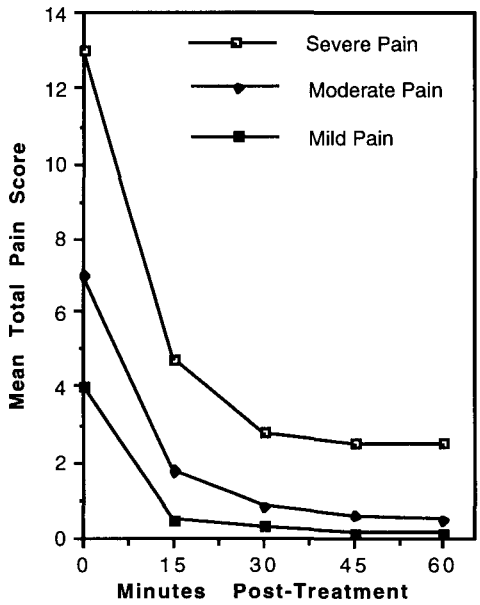


Figure 8.3A Mean total scores for all pain parameters for horses with colic that received butorphanol. (From Gingerich, D.A., Rourke, J.E., Chatfield, R.C., and Strom, P.W. 1985. Butorphanol tartrate: a new analgesic to relieve the pain of equine colic. *Vet. Med.* 80[8]:72-77. With permission.)

treatment ranged from less than 30 minutes to 75 hours (mean duration, 6.9 hours). Results (pre-treatment and post-treatment heart and respiratory rates) were analyzed using Student's t-test of paired observations.

3. Results and Conclusions

Horses were evaluated for signs of pain and discomfort during the pretreatment control period and at 15-minute intervals for 1 hour after treatment. Clinical signs (sweating, kicking, pawing and head and body movements) were each scored on a scale of 0 (none) to 4 (excessive) and summed to represent a "pain intensity score," with a range of possible values from 0 to 16. Heart and respiratory rates were also used to monitor the treatment response. The performance of the analgesic was rated according to the following criteria:

Excellent - marked analgesic effect for a period long enough to allow alleviation of the intestinal problem by specific therapy.
Good - noticeable analgesic effect but minor indications of pain still present.
Fair - only a small analgesic effect.
Poor - no analgesic effect.

The results are depicted in Figures 8.3 A-C. The 13 equine practitioners who conducted this trial rated the analgesic effect as excellent or good in 92% of the 206 cases in the study. The authors conclude that butorphanol is a safe and effective analgesic for the relief of abdominal pain in horses.

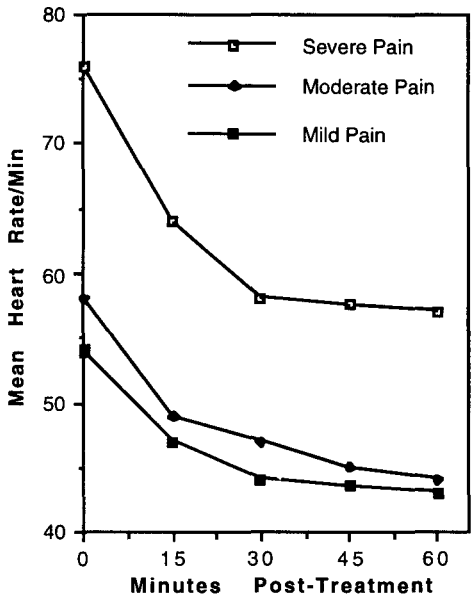


Figure 8.3B Pretreatment and post-treatment mean heart rates in horses with colic that received butorphanol. (From Gingerich et al, 1985. With permission.)

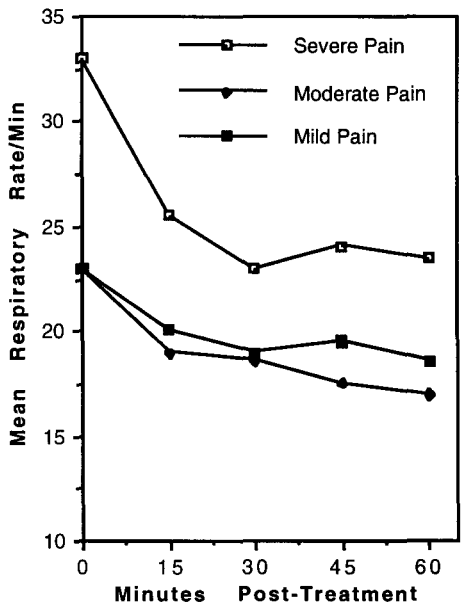


Figure 8.3C Pretreatment and post-treatment mean respiratory rates in horses with colic that received butorphanol. (From Gingerich et al, 1985. With permission.)

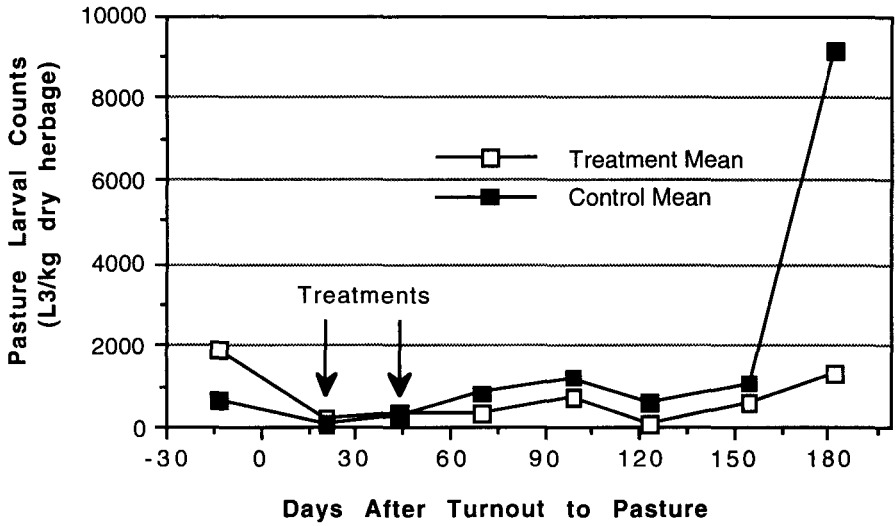


Figure 8.4 Efficacy of albendazole prophylactically – pasture larval counts. (From Herd, R.P. and Heider, L.E. 1985. Control of nematodes in dairy heifers by prophylactic treatments with albendazole in the spring. *J.A.V.M.A.* 186:1071-1074. With permission.)

4. Comments

Due to the broad range of clinical severity among horses, stratification (clinical staging) based on severity of colic increased the likelihood of detecting patient responses to treatment.

B. CASE 2: PROPHYLACTIC WORMINGS (HERD AND HEIDER, 1985)

1. Background

Dairy replacement heifers are particularly susceptible to nematode infection during their first grazing season. They frequently suffer clinical and/or subclinical infections and their high fecal egg counts are a serious source of pasture contamination and infectivity.

2. Study Design

A clinical trial was conducted to evaluate the effect of prophylactic anthelmintic treatments with albendazole 3 and 6 weeks after turnout to spring pasture around the first of May. Heifers were assigned to either lightweight (n = 12) or heavyweight (n = 10) groups (i.e., they were *blocked* by weight). Within each group, they were paired by initial weight, and one member of each pair was randomly assigned to the treatment group. The other member of each pair was an untreated control (i.e., assignment to treatment group was by *stratified randomization*). Each of the four resulting groups grazed separate, contaminated pastures until winter housing at the end of October. Weight gains were compared using analysis of variance.

3. Results and Conclusions

The strategy resulted in significant weight gain differences between treated and control *lightweight* heifers, and hastened the time at which first breeding was possible. There was no significant difference in weight gain between heavyweight groups. The study demonstrated the beneficial effects of the strategy in reducing concentrations of infective larvae on pastures. There was a sevenfold difference in larval densities between treatment and control pastures by the end of the grazing season (Figure 8.4). The authors recommend treating all dairy heifers in northern regions during their first spring at pasture.

4. Comments

Blocking and pairing (by weight) increased the likelihood of detecting differences among treatment groups. Although the investigators were not blinded with regard to treatment groups, egg per gram counts are objective measures not likely to be affected by prior knowledge of treatment status.

C. CASE 3: SURGICAL TREATMENT OF OSTEOCHONDROSIS (SMITH ET AL, 1985)

1. Background

Osteochondrosis is a disease that affects cartilage formation in young, rapidly growing animals of many species. Cartilage flap separation can develop in a variety of joint locations, resulting in an inflammatory response termed osteochondritis dissecans (OCD). Cartilage flap removal has been advocated to alleviate clinical signs of OCD of the talus, but reports of the benefit of this procedure are conflicting.

2. Study Design

OCD of the medial aspect of the talus was diagnosed in 17 joints in 11 dogs. Arthrotomy for flap removal and curettage was performed on 11 joints; six joints did not receive surgery. After a mean period of 34 months following diagnosis, the dogs were examined clinically and the affected joints were radiographed. Physical examinations and radiographic interpretations were performed independently by two clinicians; one was unaware of the medical history of each dog, except that OCD of the talus had been diagnosed. The degree of lameness, range of motion and stability of the tarsocrural joint were graded for each limb.

3. Results and Conclusions

The authors were not able to differentiate dogs that were surgically treated from those that were not. They concluded that the recommended surgical procedures did not modify progression of osteoarthritic changes.

4. Comments

The designation of the joint as the experimental unit increased the size of comparison groups. However, the actual number of dogs in the study was relatively small, raising a question as to whether there was a chance of failing to detect improvement if it occurred.

D. CASE 4: SURGICAL ASEPSIS (VASSEUR ET AL, 1985)

1. Background

Excessive and indiscriminate use of antibiotics is believed to contribute to the development of superinfections, resistant bacterial species and nosocomial infections. It has been established in human surgical patients that antibiotic prophylaxis is not routinely necessary in clean surgical procedures. Controlled studies of veterinary surgical patients have not been reported.

2. Study Design

A total of 121 dogs and seven cats were assigned randomly to be given either ampicillin (group 1) or a placebo consisting of normal saline (group 2) by the pharmacy of a VMTH. All surgical procedures (21 different operations) were classified as clean and performed by one of two surgeons participating in the study. The surgeons were responsible for evaluation of the surgical wounds, but they were unaware of which medication had been given to the patients until after the study was concluded and the incidence of postsurgical infections determined. Results (number of infections in the two groups) were compared using Fischer's exact test for a 2 by 2 table.

3. Results and Conclusions

Wound infection developed in one of the dogs given ampicillin and in none of the animals given placebo. The difference in infection rates between the two groups was not statistically significant. The authors concluded that antibiotic administration is not indicated for routine, clean surgical procedures in dogs and cats.

4. Comments

This is a nice example of a blinded study design. It is not clear why the relatively small number of cats was included in the study.

E. CASE 5: FLEABITE ALLERGIC DERMATITIS (KUNKLE AND MILCARSKY, 1985)

1. Background

Fleabite allergic dermatitis is the most common hypersensitivity skin disease of dogs and cats. Current treatment is symptomatic, consisting primarily of flea control and corticosteroids. Several investigators have reported success with flea antigen hyposensitization, but the trials were not controlled.

2. Study Design

A study was conducted to evaluate intradermal (ID) and subcutaneous (SC) administration of flea antigen to cats with signs of fleabite allergic dermatitis and living in a geographic area where flea exposure was likely to be continuous. A total of 25 adult cats were recruited from (1) a VMTH, (2) local veterinary practices and (3) local cat owners. Diagnosis of fleabite allergic dermatitis was confirmed by ID skin testing with whole flea extract. An explanation of the double-blind approach was given to all owners before their consent was obtained. Seven control cats were given saline solution ID (n = 3) or SC (n = 4) and the remaining 18 cats were given flea antigen (Greer Laboratories, Lenoir, NC) ID (n = 8) or SC (n = 10). Injections were given weekly for 20 consecutive weeks.

Owners were instructed to make no changes in their present flea control program for the duration of the study. Use of corticosteroids was discouraged, but permitted on humanitarian grounds when deemed necessary by the owner and primary investigator. Owners were informed that if the flea antigen were found to be efficacious, all cats receiving the carrier vehicle would be given the opportunity to subsequently cross over into a flea antigen-treated group.

The clinical severity of each cat's condition was graded regularly by one investigator, who was unaware of the group to which each was assigned. A separate scale was used by the owners, who were also unaware of the treatment groups. A statistical model was used to evaluate the investigator's and owners' scores, and the degree of correlation compared with Kendall's Tau test.

3. Results and Conclusions

Investigator and owner scores are depicted in Figures 8.5 A and B. Two of the cats, which had suffered from fleabite dermatitis for 1 1/2 and 5 years, respectively, apparently became desensitized naturally. Supplemental medication (corticosteroids) was given to seven cats at some point during the trial. In all groups, there was little variation in scores from 1 month to the next, as assessed by the owner or the investigator. The authors concluded that flea antigen injections cannot be recommended for therapy of fleabite allergic dermatitis in the cat.

4. Comments

This is a nice example of a double-blind clinical trial. It is interesting that both clinicians and owners tended to rank outcomes among comparison groups the same.

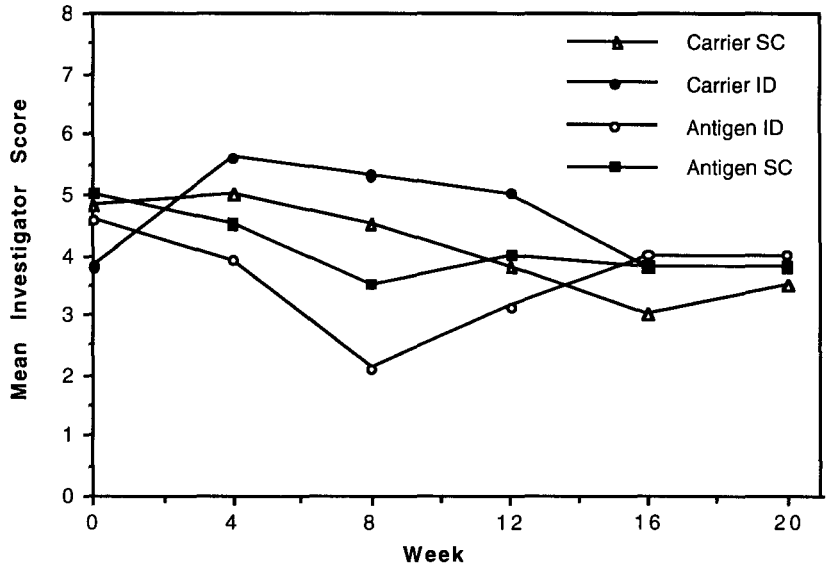


Figure 8.5A Mean *investigator scores* for four groups of cats treated with either flea antigen or placebo for fleabite allergic dermatitis over a 20-week period. SC = subcutaneous; ID = intradermal. (From Kunkle, G.A. and Milcarsky, J. 1985. Double-blind flea hyposensitization trial in cats. *J.A.V.M.A.* 186:677-680. With permission.)

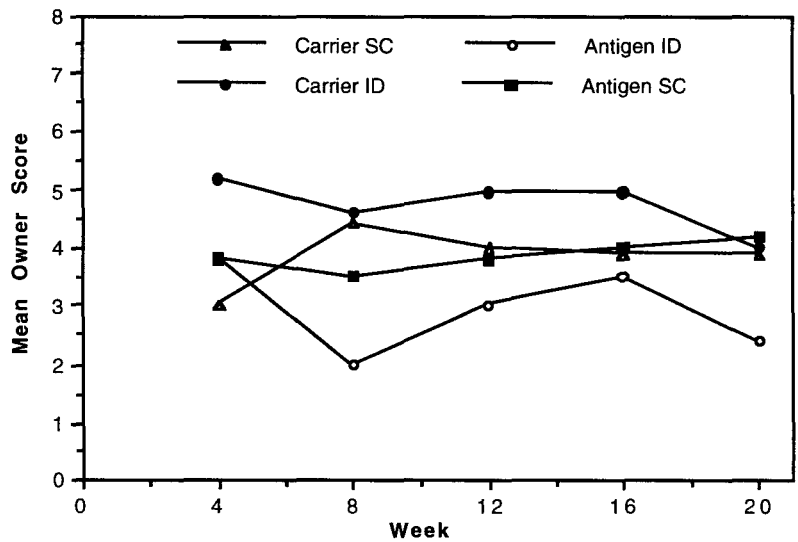


Figure 8.5B Mean *owner scores* for four groups of cats treated with either flea antigen or placebo for fleabite allergic dermatitis over a 20-week period. SC = subcutaneous; ID = intradermal. (From Kunkle, G.A. and Milcarsky, J. 1985. Double-blind flea hyposensitization trial in cats. *J.A.V.M.A.* 186:677-680. With permission.)

Table 8.2 Evaluation of case studies

Criteria	*Case 1 Equine Colic*	*Case 2 Prophy-lactic Wormings*	*Case 3 Osteo-chodrosis Surgery*	*Case 4 Surgical Asepsis*	*Case 5 Fleabite Allergy*
1. Is the case definition sufficiently explicit to exclude similar conditions?	Yes	Yes	Yes	Yes	Yes
2. Is a comparison group explicitly identified?	No	Yes		Yes	Yes
3. Are patients in each experimental group selected from the same time?		Yes			
The same place?	No	Yes	Yes	Yes	No
4. Are patients allocated to experimental groups without bias?	No	Yes	No	Yes	
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?	Yes	Yes	Yes	Yes	No
6. Is the outcome assessed without regard to treatment status?	No	Yes	Yes	Yes	Yes
7. Was the "significance" of results determined statistically?	Yes	Yes	No	Yes	Yes

V. 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.

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.

EXAMPLE: Hoskins et al (1985) evaluated the case records for 416 heartworm-infected dogs for complications following treatment with thiacetarsamide sodium (Caparsolate). Complications occurred in 26.2% of dogs and were most frequently seen 5 to 9 days following therapy. Frequency of selected complications ranged from 95.4% (increased lung sounds) to 0.9% (disseminated intravascular coagulopathy). There were no statistically significant differences between the age, sex, body size or breed of dogs that experienced complications and those that did not. However, 56 of 65 breeds were represented by six or fewer patients and had to be excluded from the statistical analysis.

VI. CLINICAL TRIALS IN PRACTICE

Randomized controlled trials are the best available means of assessing the value of treatment. Because of many practical difficulties with randomized controlled clinical trials, the majority of therapeutic questions are answered by other means, particularly uncontrolled and non-randomized trials. The need to administer some sort of treatment is largely responsible for the large percentage of case reports and uncontrolled clinical trials (see Figure 1.1).

VII. SUMMARY

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 supports the authors' conclusions, but also whether the study design was appropriate for the question being asked.

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 individuals (or their owners) who adhere to the prescribed treatment regimen. Thus an efficacious treatment could be ineffective due to poor compliance.

The clinical trial is a cohort study 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.

Virtually any parameter can be used to measure and express the outcome of a clinical trial. In veterinary medicine the outcome is often expressed in terms of productivity or economic benefit, rather than the health status of individuals.

Many factors can affect the outcome of cohort studies of risk, prognosis and treatment. Among the most important are:

1. Is the case definition sufficiently explicit to exclude similar conditions?
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 treated 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?

Chapter 9

STATISTICAL SIGNIFICANCE

I. 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 (Shott, 1985). The results of a recent review of statistical test usage in articles published in a veterinary practice journal are summarized in Table 9.1 (Smith, 1988).

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. 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 size 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 statistical routines are built into hand-held calculators, while others are available on mainframe computers or as microcomputer 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.

II. INTERPRETATION OF STATISTICAL ANALYSES

Many of the rules that apply to the interpretation of statistical tests in clinical epidemiology 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 popula-

Table 9.1 Statistical tests used in 32 of 146 articles surveyed in the Journal of the American Veterinary Medical Association, Vol. 189 (July to December, 1986)

Statistical Test/Distribution	*No. of Articles*	*% of Total*
Student's t-test	11	28.9
Chi square	9	23.7
Analysis of variance	6	15.8
Least squares regression	6	15.8
Binomial distribution	2	5.3
Normal distribution	1	2.6
Multiple logistic regression	1	2.6
Nonparametric variance analysis	1	2.6
Wilcoxon signed rank test	1	2.6

From Smith, R.D. 1988. Veterinary clinical research: a survey of study designs and clinical issues appearing in a practice journal. *Journal of Veterinary Medical Education* 15(1):2-7. With permission.

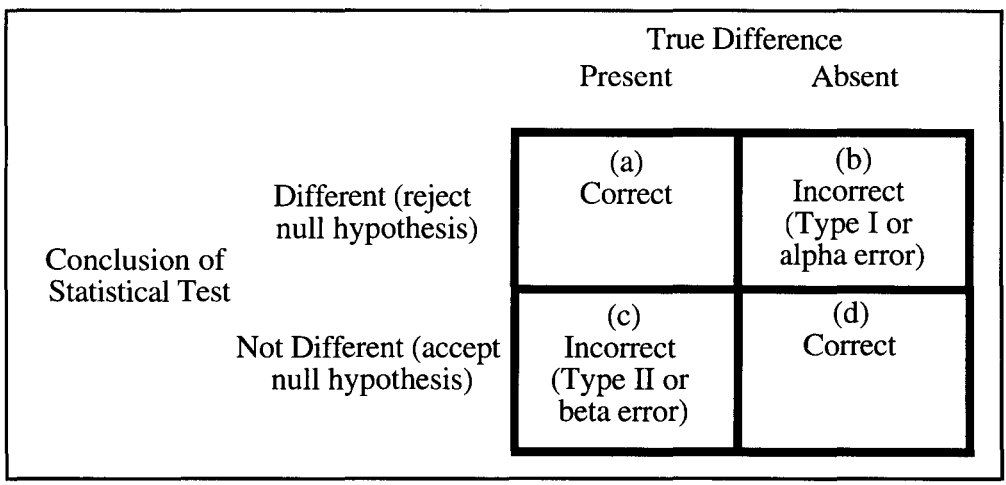


Figure 9.1 The relationship between the results of a statistical test and the true difference between possible outcomes.

tion, 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 the outcomes were really different (cell a) or they were 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.

A. CONCLUDING A DIFFERENCE EXISTS

1. The Null Hypothesis

Statistical tests reported in the medical literature are usually used to disprove the *null hypothesis*, e.g., the assumption that no difference exists between groups. If differences are detected, they are reported with the corresponding P value, which expresses the likelihood that the observed differences could have arisen by chance alone. This P value is sometimes referred to as "P_a" to distinguish it from beta error.

2. Statistical Significance

A P value is usually considered to be *statistically significant* if it falls below 0.05, e.g., 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.

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. 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 (see Figure 9.2 for an example of a statistically significant *association* that is not clinically significant).

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.

3. Confidence Intervals

The confidence interval 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. 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.

EXAMPLE: The American Veterinary Medical Association (AVMA) conducted an economic survey of U.S. veterinarians in the spring of 1992 (Wise, 1993). The purpose of the survey was to secure accurate data on veterinarians' earnings in private and nonprivate practice. Individuals were selected randomly from AVMA's computerized records of 45,651 nonretired member and nonmember veterinarians. A total of 3909 (40%) of 9799 veterinarians surveyed responded. Median and mean incomes were estimated and compared among six practice and six nonpractice types (Table 9.2). Estimates derived from sample surveys are subject to sampling errors (bias) that arise because observations are made on only a portion of the total population. Therefore, 95% confidence intervals were used to draw inferences on the magnitude of differences of mean salaries among employment categories. We can be 95% sure that the estimated mean income plus or minus 1.96 SDs of the mean (standard error or SEM) will encompass the true, but unknown, population mean income for each group. Note that the median income is consistently lower than the mean due to positively skewed salary distributions. Because the mean income is influenced by extreme values at the high end of the income distribution, the median estimate often is considered a more meaningful estimate of the central income level of a population.

Table 9.2 Median and mean 1991 incomes of U.S. veterinarians in private and nonprivate practice, ranked according to mean income

Employment Category	*Estimated Median Income ($)*	*Estimated Mean Income ($)*	*95% Confidence Interval of True Mean Income (%)*
Uniformed services	50,500	50,658	$48.669 \leq \mu \leq 52,647$
Mixed animal	41,725	50,968	$47,780 \leq \mu \leq 54,156$
State or local government	50,500	52,442	$50,052 \leq \mu \leq 54,832$
Federal government	50,500	54,277	$51,701 \leq \mu \leq 56,852$
Large animal predominant	45,736	60,027	$55,502 \leq \mu \leq 64,552$
Small animal predominant	45,100	61,479	$56,398 \leq \mu \leq 66,560$
Other/not-for-profit organization	59,500	63,676	$57,273 \leq \mu \leq 70,078$
Large animal exclusive	53,500	63,678	$57,989 \leq \mu \leq 69,368$
Small animal exclusive	47,500	65,316	$60,277 \leq \mu \leq 70,354$
College or university	65,500	67,265	$62,922 \leq \mu \leq 71,609$
Equine predominant	50,500	68,918	$63,062 \leq \mu \leq 74,774$

From Wise, J.K. 1993. 1991 professional incomes of US veterinarians. *J.A.V.M.A.* 202:210-212. With permission.

4. Confidence Interval for a Rate or Proportion

The confidence intervals reported in Table 9.2 were estimated by using the individual values (incomes) reported by survey respondents to calculate the mean, variance, and standard deviation of income levels. It is also possible to estimate the confidence interval for a proportion such as the prevalence of disease by using the binomial distribution (Huntsberger and Billingsley, 1973). In this approach the disease prevalence value is considered to be the mean. The variance of disease prevalence = [p(1 - p)/n], where n = sample size and p = proportion of diseased individuals. The standard deviation of disease prevalence = the square root of the variance.

For example in Table 5.2 the prevalence (p) of *M. paratuberculosis* among Illinois cattle (n = 171) was 1.2%.

$$\text{The variance of the disease prevalence} = \frac{0.012 * .988}{171} = .0000693$$

The standard deviation of the disease prevalence (square root of the variance) = 0.00832 or $\approx$ 0.8%, which is consistent with the estimate reported by the investigators. The 95% confidence interval for the prevalence of *M. paratuberculosis* would be 1.2% +/- 1.96(0.832%), or -0.4% to 2.8%. The fact that there is a chance that *M paratuberculosis* prevalence could be less than 0%, even though the organism was isolated from ileocecal lymph nodes, results from the fact that the binomial distribution of proportions is not symmetrical around the mean, except for the special case where p = 0.50.

5. One-Tailed Versus Two-Tailed Tests

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, e.g., 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. Refer to Figures 2.7 and 2.8 for a comparison of one- and two-tailed cutoffs.

B. CONCLUDING A DIFFERENCE DOES NOT EXIST

1. Statistical Significance

By default, P values ≥ 0.05 imply that no difference between outcomes or treatment groups exist. This does not exclude the chance, however, 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 P_b.

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 disease eradication programs that rely on diagnostic tests to identify infected animals or herds, e.g., distinguish them from uninfected herds, even when the number of infected animals is low. Sample size is discussed further in the following sections.

C. CONCLUDING AN ASSOCIATION EXISTS

1. Agreement Between Tests

As stated in Chapter 3 (Evaluation of Diagnostic Tests), concordance is the proportion of all test results on which two or more different tests agree. The level of agreement is frequently expressed as the kappa (k) statistic, defined as the proportion of potential agreement beyond chance exhibited by two or more tests. Expected agreement by chance alone is calculated by the method of marginal cross products. The value of kappa ranges from -1.0 (perfect disagreement) through 0.0 (chance agreement only) to +1.0 (perfect agreement). By convention, kappa values of 0.0 - 0.2 = slight, 0.2 - 0.4 = fair, 0.4 - 0.6 = moderate, 0.6 - 0.8 = substantial, and 0.8 - 1.0 = almost perfect agreement between tests (Sackett, 1992).

To illustrate how the kappa statistic is used, let us compare an ELISA test for circulating heartworm (*Dirofilaria immitis*) antigen with the modified Knott's test for circulating microfilariae (Figure 9.2; Courtney et al, 1990). In this study there were 341 heartworm-infected and 206 heartworm-uninfected dogs. Infection status (gold standard) was determined at necropsy. Although none of the uninfected dogs harbored adult *D. immitis*, 22 had circulating microfilariae of *Dipetalonema reconditum* and one had circulating microfilariae of both *D. immitis* and *D. reconditum*.

Test concordance was 82% [(201 + 247) ÷ 547]. On the basis of column and row totals we would expect the two tests to agree 49% of the time by chance alone, and the remaining potential agreement beyond chance would therefore be 100% - 49% or 51%. The observed agreement beyond chance was 82% - 49% or 33%, yielding a value for kappa of 0.65. In this case (k = 0.65), there was "substantial" agreement between the Knott's and ELISA tests.

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. In this study 41% (140 out of 341) of heartworm infections were occult and undetectable by the Knott's test. The ELISA

KNOTT'S TEST

ELISA	Positive	Negative	
Positive	(a) 201 (110)	(b) 98	(a + b) 299
Negative	(c) 1	(d) 247 (156)	(c + d) 248
	(a + c) 202	(b + d) 345	(a + b + c + d) 547

Figure 9.2 Two by two table comparing concordance of Knott's and ELISA test results for *Dirofilaria immitis* infection in dogs. Numbers in parentheses are expected values based on the method of marginal cross products. (Source of data: Courtney, C.H., Zeng, Q.Y., and Tonelli, Q. 1990. Sensitivity and specificity of the CITE heartworm antigen test and a comparison with the DiroChek heartworm antigen test. *J. Am. Anim. Hosp. Assoc.* 26:623-628.)

Observed agreement (concordance) =

$$\frac{a + d}{a + b + c + d} = \frac{(\text{observed a}) + (\text{observed d})}{a + b + c + d} = \frac{(201 + 247)}{547} = 82\%$$

$$\text{Expected (chance) agreement for cell a} = \frac{(a + b) \times (a + c)}{a + b + c + d} = \frac{(299 \times 202)}{547} = 110$$

$$\text{Expected (chance) agreement for cell d} = \frac{(c + d) \times (b + d)}{a + b + c + d} = \frac{(248 \times 345)}{547} = 156$$

$$\text{Expected (chance) agreement overall} = \frac{(\text{expected a}) + (\text{expected d})}{a + b + c + d} = \frac{(110 + 156)}{547} = 49\%$$

Agreement beyond chance (kappa) =

$$\frac{\text{observed agreement - expected agreement}}{100\% - \text{expected agreement}} = \frac{82\% - 49\%}{100\% - 49\%} = \frac{33\%}{51\%} = 0.65$$

test detected 65% (91) of these, which accounts for most of the ELISA-positive/Knott's-negative test results in cell "b."

2. Linear Association Between Two Variables

Statistics are also used to describe the degree of association between variables. The *correlation coefficient*, r (formally known as the *Pearson product-moment coefficient of correlation*, or the *Pearson r*), is a measure of the strength and direction of a linear association between two

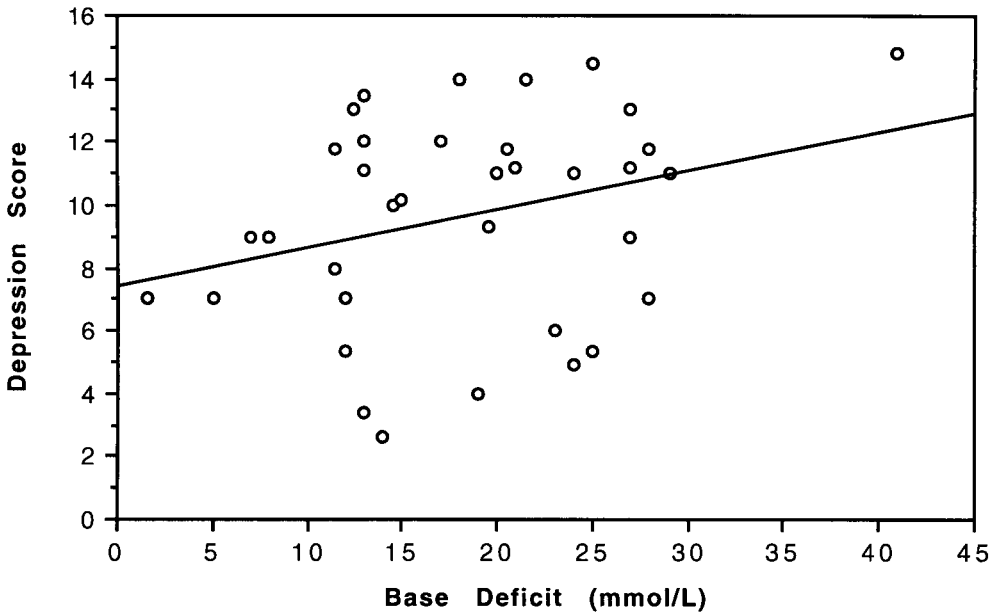


Figure 9.3 Relationship between 0-hour depression score and base deficit in 36 dehydrated diarrheic calves. (From Kasari, T.R. and Naylor, J.M. 1985. Clinical evaluation of sodium bicarbonate, sodium L-lactate, and sodium acetate for the treatment of acidosis in diarrheic calves. *J.A.V.M.A.* 187:392-397. With permission.)

interval-level variables (Sharp, 1979). 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 is a measure of closeness of fit of the data to the linear regression line. The value for r^2 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.

The *Spearman rank coefficient*, or *Spearman rho (p)*, is the counterpart of the Pearson coefficient of correlation (r) for ordinal data. It is a nonparametric measure (see following) for use with data that are either reduced to ranks or collected in the form of ranks. The Spearman rho, like the Pearson coefficient of correlation, yields a value from -1 to 1, and it is interpreted in the same way (Sharp, 1979).

EXAMPLE: Thirty-six dehydrated diarrheic neonatal calves were used to study the correlation of clinical condition (staging) with acid-base (base deficit) status, using a scoring system for depression (Kasari and Naylor, 1985). The hypothesized association between these two variables is depicted as a scattergram in Figure 9.3. There was a statistically significant ($r = 0.30$, $P < 0.05$) linear relationship between depression score and base deficit, but this relationship accounted for less than 10% ($r^2 = 0.09$) of the individual variation in acid-base status.

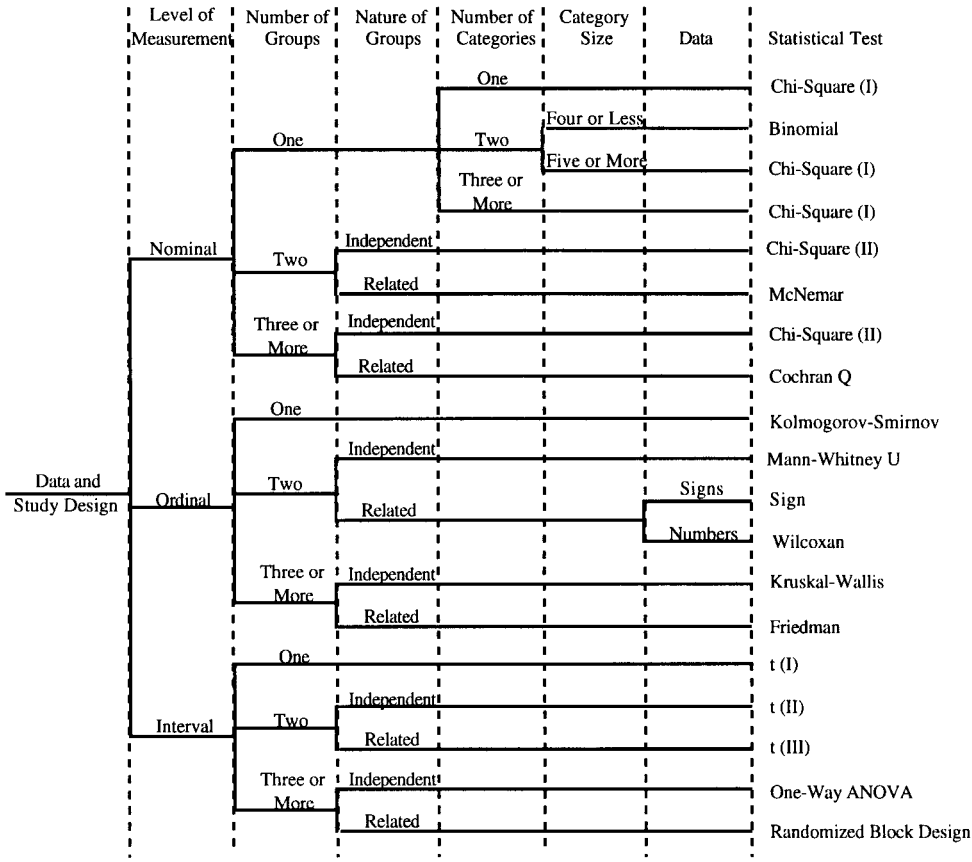


Figure 9.4 Tree diagram for selection of an appropriate statistical test depending upon characteristics of the study design and data to be analyzed. (Adapted from Sharp, V.F. 1979. *Statistics for the Social Sciences.* Little, Brown & Co., Boston, MA. 381 pp. With permission.)

Thus, the clinical scoring system was of limited use in predicting the total bicarbonate ion losses in individual dehydrated diarrheic calves.

III. THE SELECTION OF AN APPROPRIATE STATISTICAL TEST

All of the common statistical tests are used to estimate the probability of an alpha error, e.g., 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_a may be misleading.

In research there are many different statistical tests of significance. 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 decide which statistical test is appropriate for a particular study design.

For the uninitiated (most of us), the choice of an appropriate statistical test is not intuitively obvious. The tree diagram in Figure 9.4 provides guidelines for 15 of the most widely used statistical tests (Sharp, 1979). It takes into account the major requirements of each sta-

tistical test, which serve as directions for determining the appropriate test. Relevant questions for each branch of the tree follow.

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_a may be misleading.

A. LEVEL OF MEASUREMENT

What is the level of measurement: nominal, ordinal or interval? *Nominal data* is used to categorize objects, individuals, conditions, etc. without ranking, as breed, sex or blood line. *Ordinal data* is ranked but does not fall on a uniform scale. Terms such as "light," "moderate" and "heavy" are used to describe ordinal data. *Interval data* is 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.

B. 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*.

C. 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, 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, sex or breed before randomly assigning them 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 regression analysis where variables other than the one under consideration may influence the outcome.

D. 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 can be separated into three categories based on sex: male, female or neutered.

E. CATEGORY SIZE

How many individuals or objects are in each of your categories? This question also refers only to nominal data.

F. DATA

How do you plan to use your data? This question only applies to ordinal data divided into

Table 9.3 Nonparametric and parametric statistical tests listed in Figure 9.3

Nonparametric tests
Binomial (test of proportion)
Chi-square (I) (goodness of fit test of observed versus expected frequencies)
Chi-square (II) (contingency table analysis)
McNemar
Cochran Q
Kolmogorov-Smirnov
Mann-Whitney U
Sign
Wilcoxan
Kruskal-Wallis
Friedman
Spearman rho (*p*)*
Parametric tests
t (I) (compares sample with population mean)
t (II) (unpaired t-test)
t (III) (paired t-test)
One-way analysis of variance
Randomized blocks design (two-way analysis of variance)
Pearson r*

*Spearman rho and Pearson r are measures of the degree of correlation between two variables. They do not appear in Figure 9.3.

From Sharp, V.F. 1979. *Statistics for the Social Sciences.* Little, Brown & Co., Boston, MA. 381 pp. With permission.

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).

IV. PARAMETRIC AND NONPARAMETRIC TESTS

Statistical tests are referred to as either *parametric* or *nonparametric*. When choosing a statistical test using the tree in Figure 9.4, we are also making a choice between a parametric or nonparametric test. Statistical tests appearing in the tree are organized as nonparametric or parametric in Table 9.3 (Sharp, 1979).

Parametric tests are more powerful than nonparametric tests, e.g., 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 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 must be used for very small sample sizes, e.g., six or fewer (Sharp, 1979).

V. USING A TREE DIAGRAM TO SELECT A STATISTICAL TEST

The use of the tree diagram can be demonstrated using some of the case studies from Chapter 8 (see section on clinical trials and Table 8.2).

Case 1 – Treatment of equine colic: The investigators evaluated the effect of an analgesic on a single pretreatment and post-treatment pain intensity score taken in each patient. Scores were expressed numerically as *interval-level variables*. There were only *two groups*, pretreatment and post-treatment, and because of repeated measures the *groups were related*. The authors correctly chose to use t (III), the *paired t-test*, to compare the results of pretreatment and post-treatment pain intensity scores.

Case 2 – Prophylactic wormings: The investigators measured the effect of prophylactic wormings on weight gains among four groups of cattle that were formed by stratification and pairing according to weight, followed by random assignment to treatment or control groups. Weight gain is an *interval-level variable*. There were *four groups*, treated and control heavy and lightweight heifers, which were *related* because of pairing. The authors correctly chose a *two-way analysis of variance* (for randomized blocks design). The experimental design dictated the type of statistical test to be performed.

Case 4 – Surgical asepsis: The investigators compared the number of wound infections in "clean" canine and feline surgeries where prophylactic antibiotics were or were not used. The outcome data (infection present or absent) is *nominal* and is distributed over *two groups* of patients – antibiotics given or not given. The *groups were independent*. The authors correctly chose *Fisher's exact test* (a modification of Chi-square II) to analyze their data.

It is intuitively obvious that the more subjects that are entered into a study, the more faith we can have that differences among groups are not due to random variation. The question is, how many subjects are enough?

VI. SAMPLE SIZE

It is intuitively obvious that the more subjects that are entered into a study, the more faith 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_a and (5) P_b. Three common situations in which sample size must be considered follow.

A. 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 or not a particular disease is present in the herd. This is a common concern in disease eradication or control programs, such as Illinois' swine pseudorabies eradication program. 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: Consider a herd of pigs in which 10% are infected with the pseudorabies virus and have detectable serum antibody. If a serum sample is drawn from one randomly selected animal in the herd, the probability that it will come from a pseudorabies-free animal is 0.90. Thus, P_b is 0.90 and we have a 90% chance of failing to detect infection in the herd. If two animals are sampled, then the chance that both samples were drawn from negative animals is 0.90 x 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 turned around 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 infected with pseudorabies virus.

The astute reader will have noticed that the previous formula is true only for very large herd sizes. For example, if the swine herd consisted of 29 animals or less, and all were tested, we would be more than 95% sure that at least one of the sampled animals was infected with pseudorabies. 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.

B. MINIMUM SAMPLE SIZE FOR ESTIMATING A RATE 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 used 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) (Cochran, 1977, p 75).

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 not included 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 when sampling from a *finite population* (N) the following conversion (Cochran, 1977, p 76) can be made:

$$n_{fin} = \frac{n_{inf}}{1 + (n_{inf} - 1)/N}$$

C. MINIMUM SAMPLE SIZE TO DETECT DIFFERENCES AMONG GROUPS IN STUDIES OF RISK, PROGNOSIS AND TREATMENT

As indicated previously, a variety of statistical tests is 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.

VII. MULTIPLE COMPARISONS

Some studies, called "*hypothesis testing*," are designed to evaluate the effect of one variable (as a risk factor, prognostic factor or treatment) on 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 generating*" (aka "data dredging," "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.

EXAMPLE: In the study of causes of death in veterinarians (see Table 6.4), the authors initially compared veterinarians, as a group, with nonveterinarians. This led to the identification of increased risks of death from some diseases and reduced risk for others. The investigators then broke the group of veterinarians down into subgroups, based on their specialties. When this was done, some interesting risks emerged for veterinarians in certain specialties, but the numbers were too small to be statistically significant.

If enough comparisons are made, the more likely that at least one will be statistically significant, irrespective of the true state of affairs.

A second problem in making multiple comparisons is similar to the problem encountered in parallel testing – if enough comparisons are made, the more likely that at least one will be statistically significant, irrespective of the true state of affairs. Consequently, results derived

from multiple comparisons should be considered as hypotheses to be tested in follow-up studies.

VIII. SUMMARY

Statistical analyses, once a rarity in medical journals, are now routinely encountered in the medical literature, and veterinary journals are no exception. Such analyses often have immense practical importance, since research results are frequently the basis for decisions about patient care.

Many of the rules that apply to the interpretation of statistical tests in clinical epidemiology 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. There are thus four possible outcomes of statistical tests – two are correct and two are incorrect. Alpha or Type I error results when we conclude that outcomes were different when, in fact, they were not. Alpha error is analogous to the false-positive result of diagnostic tests. Beta or Type II error occurs when we conclude that outcomes were not different when, in fact, they were. Beta error is analogous to the false-negative result of diagnostic tests.

Statistical tests reported in the medical literature are usually used to disprove the null hypothesis, e.g., the assumption that no difference exists between groups. If differences are detected, they are reported with the corresponding P value, which expresses the likelihood that the observed differences could have arisen by chance alone. This P value is sometimes referred to as "P_a" to distinguish it from beta error. A P value is usually considered to be statistically significant if it falls below 0.05, e.g., 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. The confidence interval provides a way of expressing the range over which a value is likely to occur.

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. Two-tailed tests are more conservative, e.g., 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.

Power is the probability that a study will find a statistically significant difference when one exists. Power is analogous to diagnostic test sensitivity. P_b is the major determinant of sample size in disease eradication programs that rely on diagnostic tests to identify infected animals or herds, e.g., distinguish them from uninfected herds, even when the number of infected animals is low.

Statistics are also used to describe the degree of association between variables. The level of agreement between two or more test results (when expressed as categorical variables) is frequently expressed as the kappa (k) statistic, defined as the proportion of potential agreement beyond chance. The value of kappa ranges from -1.0 (perfect disagreement) through 0.0 (chance agreement only) to +1.0 (perfect agreement). The correlation coefficient, r, is a measure of the degree of linear association between two interval-level variables. 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 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 is a measure of closeness of fit of the data to the linear regression line. The value for r^2 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.

All of the common statistical tests are used to estimate the probability of an alpha error, e.g., 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_a may be misleading. Among the considerations in choosing a statistical test are (1) whether the data are nominal, ordinal or interval, (2) the number of groups being compared, (3) whether the groups are independent or related, (4) the number and size of categories (for nominal data) and (5) how we intend to compare the data (for ordinal data).

It is intuitively obvious that the more subjects that are entered into a study, the more faith we can have that differences among groups are not due to random variation. The question is how many subjects are necessary to ensure the power of anticipated or published studies? 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_a and (5) P_b. Three common situations where sample size must be considered are (1) minimum sample size for demonstrating an extreme outcome, (2) minimum sample size for estimating a rate with a specified degree of precision and (3) minimum sample size to detect differences among groups in studies of risk, prognosis and treatment.

Chapter 10

MEDICAL ECOLOGY AND OUTBREAK INVESTIGATION

I. 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 discuss the dynamics of disease in populations, e.g., *medical ecology*. We also learn how to conduct outbreak investigations using all of the concepts, tools and approaches discussed in previous chapters.

One of the things that distinguishes veterinary from human medicine is the fact that veterinarians are frequently called on 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 the disease in the population.

Practitioners are frequently called on to participate in local, state and federal disease control programs. In their role as "middlemen," 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.

II. 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 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 summarized in Table 10.1.

A. 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

Table 10.1 Issues in the epidemiology of a disease

Occurrence	What is the case definition? What is the host, spatial and temporal distribution of the disease?
Cause	What is the etiologic agent? What is its life cycle? What characteristics contribute to its pathogenicity and virulence?
Susceptibility	What factors determine the susceptibility or resistance of individuals to the disease? What conditions predispose populations to outbreaks?
Source	What is the source and reservoir mechanism of the causative agent? What are the periods of communicability?
Transmission	How is the agent spread from infected to susceptible individuals? What is the route of infection?
Cost	What is the economic impact of the disease?
Control	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?

(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 on its probable cause, source and mode of transmission.

B. 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.

C. 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 *susceptibles*, *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.

D. 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.

E. 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.

F. 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.

G. CONTROL

Ultimately the practitioner must devise a plan for the reduction of disease frequency in the population. This may be accomplished through disease prevention, control (treatment) or eradication.

III. OUTBREAK INVESTIGATION

Outbreak investigation is similar, in principle, to examination of a patient in a hospital setting. In both instances history, 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) used with individual patients. Components of an epidemiologic workup include the following:

A. 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:

(A) What are the characteristics of the clinical syndrome, e.g., the case definition?

(1) What signs were/are observed in live and dead animals?
(2) What was the incubation period?
(3) How long did signs last?
(4) What is the prognosis for diseased animals?

(B) What are the temporal, spatial and demographic patterns of disease?

(1) When did the cases occur?
(2) Where did the cases occur?
(3) What was the incidence of disease, e.g., how many animals were at risk and how many were affected?
(4) What are the characteristics of the affected and unaffected animals?
(5) How rapidly did the disease spread and what is the likely mode of transmission?
(6) Are any other domestic animal or wildlife affected; is there any concurrent human illness?

(C) What is the herd history?

(1) Describe the management and husbandry practices, including housing, feed, water.
(2) Describe disease control/hygiene practices including vaccination, parasiticides/dewormers, other treatments, vermin and pest control, and waste disposal.
(3) Describe the herd's production/disease history.
(4) Has there been contact with other domestic animals or wildlife?
(5) Has there been any animal movement or introductions recently?
(6) Have there been any health problems in adjacent herds?

(D) What is the environmental history?

(1) What has the weather been like?
(2) Describe the geographic location, e.g., topography, soil type, vegetation.
(3) Have fertilizers, herbicides, pesticides been used recently?

The answers to the above questions should help guide sample collection and the selection of appropriate diagnostic test procedures.

B. ANALYTIC PHASE (ASSESSMENT)

During this phase 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?

C. 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. Case studies are included to demonstrate how outbreak investigations are conducted.

IV. SUMMARY

A number of issues surface when considering the epidemiology of any disease. These include its cause, occurrence, source and transmission, determinants of the susceptibility of individuals and populations, the cost of the disease and measures that can be used to achieve control.

Outbreak investigation is similar, in principle, to examination of a patient in a hospital setting. In both instances history, physical and laboratory examinations are used to try to identify the cause(s) of disease at the herd or individual 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. Disease determinants are generally divided into three categories: agent, host and environmental factors.

An epidemiologic workup is similar to the clinical assessment of individual patients and includes descriptive, analytical and intervention phases. During the descriptive phase data are collected from the herd and the patterns of disease occurrence over time, space and among hosts are described. During the analytic phase the descriptive data are compared and analyzed in light of what is known about diseases on the differential list. During the intervention phase an optimal disease control plan is selected based on the best combination of immediate and long-term objectives.

Chapter 11

MEASURING AND EXPRESSING OCCURRENCE

I. 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 within a host population. This information is useful not only to gain a better appreciation of the significance of the disease, but may suggest the probable cause, source and mode of transmission of the condition.

II. 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 usually don't 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.

A. 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, may be added to the case definition.

EXAMPLE: Equine ehrlichial colitis (EEC), also known as Potomac horse fever or equine monocytic ehrlichiosis, is a recently recognized enteric disease of horses. There are a variety of clinical syndromes, ranging from fever, depression and anorexia to uncontrollable colic to severe watery diarrhea. Laminitis may also occur. Palmer et al (1986) sought to document the occurrence of the syndrome in Pennsylvania, New Jersey, New York, Ohio, Idaho and Connecticut. Potential cases were initially selected during telephone consultations with veterinarians reporting unusual enteric disease manifested as diarrhea and colic associated with colitis. In each area an increase in the occurrence of equine enteric disease, as perceived by the attending veterinarian, prompted the consultation.

The problem for the investigators was to distinguish those cases of enteric disease attributable to EEC from those which could be attributable to other causes, notably *Salmonella* sp. infections. The clinical signs of the two diseases are indistinguishable, and differentiation

Table 11.1 Epidemiologic components of the case definition used to distinguish equine ehrlichial colitis from salmonellosis

Occurrence	*Equine Ehrlichial Colitis*	*Salmonellosis*
Geographic	No concentration in any one area of a farm	Concentrated in particular areas of a farm
Temporal	Seasonal incidence from May through October; most cases occurring July through September	Occurs throughout the year
Host	Occurs in apparently "unstressed" horses, e.g., aged "retired" horses at pasture	Frequently occurs in stressed horses, foals, and weanlings

Reprinted with permission from Palmer, J.E., Whitlock, R.H., and Benson, C.E. 1986. Equine ehrlichial colitis (Potomac horse fever): recognition of the disease in Pennsylvania, New Jersey, New York, Ohio, Idaho, and Connecticut. *J.A.V.M.A.* 189:197-199.

cannot be made on the basis of clinical signs or laboratory data alone. However, the epidemiology of EEC differs from that of equine salmonellosis in several important respects (Table 11.1).

A case definition was developed from earlier reports of the disease in Montgomery County, MD, and was used to screen potential cases for follow-up. To further restrict the number of horses to be followed up, the case definition included the epidemiologic features of EEC described in Table 11.1. Infection was confirmed by indirect fluorescent antibody tests of paired sera.

Eight areas endemic for EEC were identified based on finding a fourfold or greater change (increase or decrease) in antibody titer from paired serum samples in at least one horse with clinical signs of colitis. The attack rate per farm was generally low. No attempt was made to estimate prevalence, because serum samples were available in only a few of the cases of colitis in each area. Clinical signs varied from fever and depression to severe diarrhea and laminitis. Occasionally horses developed profound ileus (hypomotility of the intestines) and severe colic. Horses on pasture, as well as those stabled, were affected.

B. 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.

EXAMPLE: A review was made of a year's records and of the relationship of animal performance and management procedures at a swine feedlot in central Kansas (Straw et al, 1985). Aspects of performance that were considered unsatisfactory included (1) slow growth rate of finishing pigs, (2) poor feed conversion, (3) high death rate (especially due to *Haemophilus* pneumonia) and (4) excessive carcass trim at the time pigs were slaughtered. During the year, there was a continuous flow of pigs into and out of the feedlot. Data were used from all groups that had been sold that year.

Analyses were performed on 38 groups containing 9988 pigs. Although overall performance was low, certain groups of pigs (defined as noncases) performed considerably better than

others (defined as cases). Comparisons between groups were made in an effort to identify management inputs (risk factors) that could be used to improve overall performance.

Factors having the greatest influence on performance were the month of entry of pigs into the feedlot, amount of injectable antibiotics used, weight of pigs on entry into the feedlot and amount of time spent in the feedlot. The investigators recommended that the producer (1) start pigs only during spring and summer months, (2) use oral antibiotic therapy if possible to avoid carcass trim at slaughter, (3) market all animals by 150 days after entry into the feedlot (regardless of weight) and (4) use a *Haemophilus* vaccine of proven efficacy.

III. 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 (Morton and Hebel, 1979).

A. HOST DISTRIBUTION

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 among the total exposed at the beginning of the outbreak (Morton and Hebel, 1979). 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.

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).

B. 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.

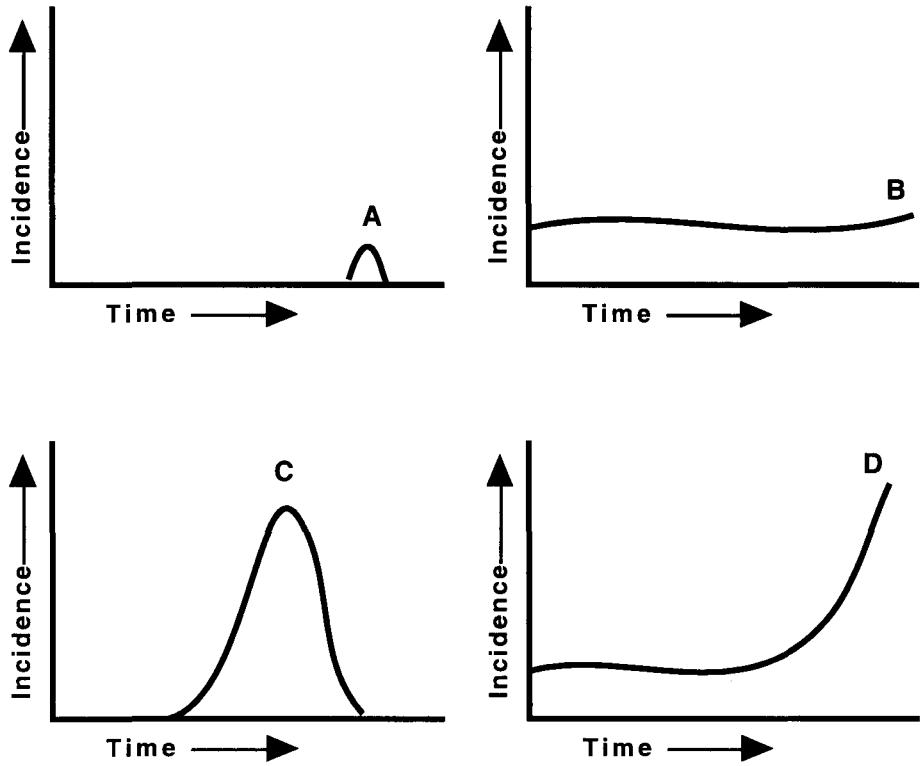


Figure 11.1 Examples of patterns of disease occurrence. (A) sporadic, (B) endemic, (C) point source epidemic and (D) propagating epidemic. (Modified with permission from Schwabe, C.W., Riemann, H.P., and Franti, C.E. 1977. *Epidemiology in Veterinary Practice.* Lea & Febiger, Philadelphia. 303 pp.)

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 brucellosis), a suitable vector (as Rocky Mountain spotted fever) or occasional contact with an environmental source, either animal (as rabies) or inanimate (as tetanus).

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, as reflected in terms used to describe the levels of occurrence of endemic disease: (1) *holoendemic*, when most animals are affected, (2) *hyperendemic*, when a high proportion of animals are affected, (3) *mesoendemic*, when a moderate proportion of animals are affected or (4) *hypoendemic*, when a relatively small proportion of animals are affected. Herd infestations with internal parasites and bovine anaplasmosis tend to occur as endemic diseases.

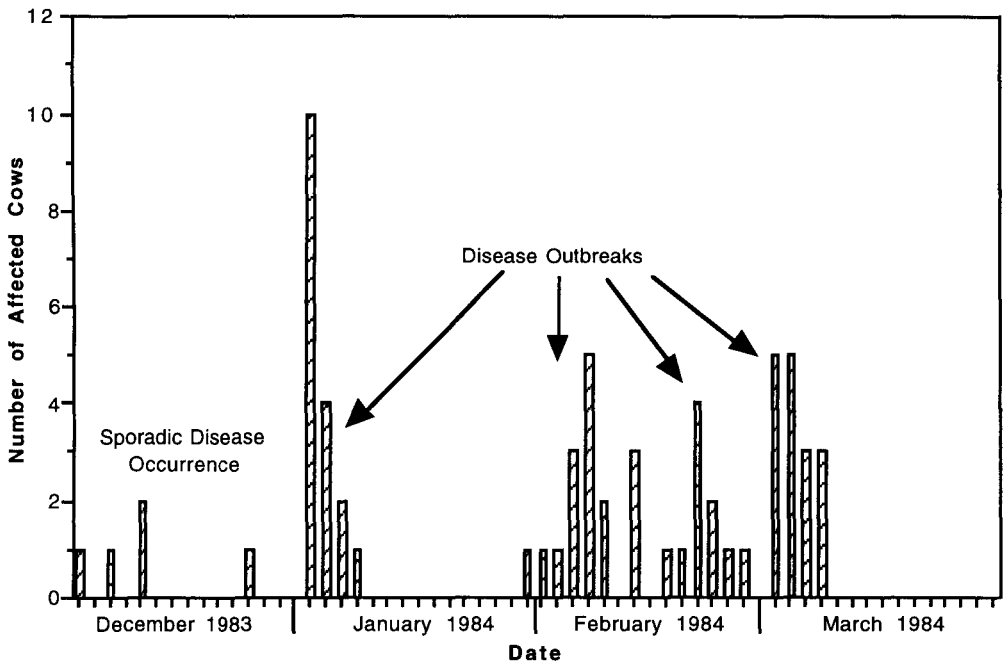


Figure 11.2 Temporal distribution of clinical mastitis treated in a herd. Sporadic incidence during December 1983 is followed by a series of epidemics from January through March 1984. (Reprinted with permission from Bowman, G.L., Hueston, W.D., Boner, G.J., Hurley, J.J., and Andreas, J.E. 1986. *Serratia liquefaciens* mastitis in a dairy herd. *J.A.V.M.A.* 189:913-915.)

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; 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. A *pandemic* is a large-scale epidemic over a wide geographic region. 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.

Stylized temporal patterns of disease occurrence are depicted in Figure 11.1, and specific examples in Figures 11.2 and 11.3. Figure 11.2 depicts sporadic occurrence (incidence during December 1983) of new cases of clinical mastitis followed by a series of epidemics. The initial sporadic cases were attributed to opportunistic infections with *Serratia liquefaciens* in teats damaged by severe cold. Subsequent epidemics were attributed to mechanical spread to other cows with damaged teats during the milking procedure (Bowman et al, 1986).

Figure 11.3 depicts an epidemic of infertility within a 940-cow dairy herd attributed to trichomoniasis (Goodger and Skirrow, 1986). Overall prevalence of infection (crude rate) during January 1985 was 10.67%, based on culture results. During the latter half of 1984 the temporal occurrence was consistent with the definition of a propagating epidemic, suggesting unabated spread of the agent to susceptible animals.

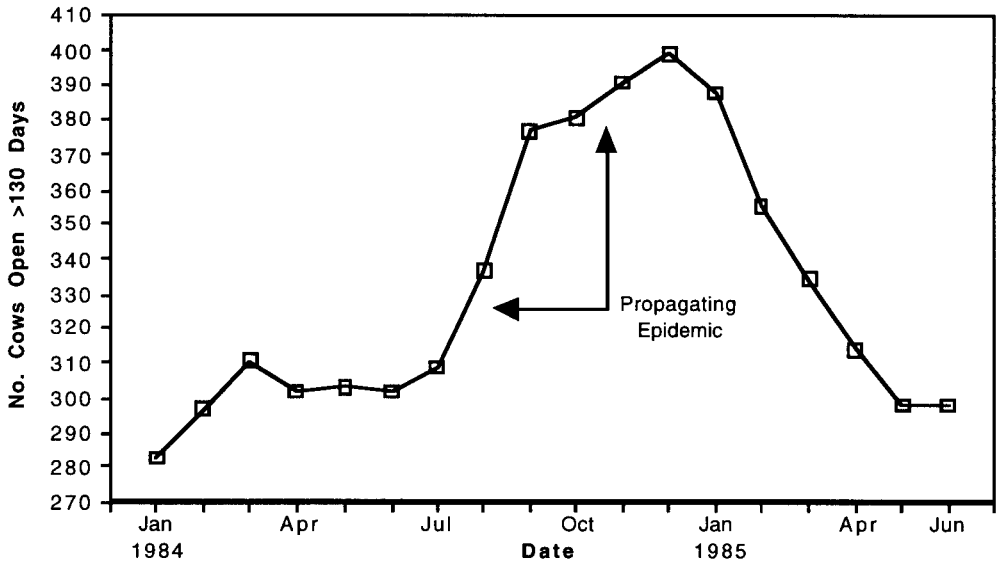


Figure 11.3 A propagating epidemic of infertility in a 940-cow dairy herd. (Reprinted with permission from Goodger, W.J. and Skirrow, S.Z. 1986. Epidemiologic and economic analyses of an unusually long epizootic of trichomoniasis in a large California dairy herd. *J.A.V.M.A.* 189:772-776.)

C. TIME SERIES ANALYSIS

Time series analysis is concerned with the detection, description and measurement of patterns or periodicities from temporal occurrence data (Schwabe et al, 1977). 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 (incidence) are influenced by one or more of the following: (1) secular trend, (2) seasonal fluctuation, (3) cyclic variation and (4) irregular variation (Carter et al, 1986).

Patterns of disease occurrence are influenced by one or more of the following: (1) secular trend, (2) seasonal fluctuation, (3) cyclic variation and (4) irregular variation.

Secular trends are overall long-term rises or declines in incidence rate that occur gradually over long periods of time. A secular trend can be identified from time series data by (1) visual observation of plotted raw data, (2) least squares regression or (3) the moving average method (Figures 11.4 and 11.5). 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 are beyond the scope of this book, but can be found in standard statistical texts. It is a standard option on statistical calculators and statistical packages for computers. A moving average is a series of data averages centered at each successive measurement point on the time scale (Schwabe et al, 1977). 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.

Seasonal fluctuations are regular changes in incidence rates with periods shorter than a year. Three-month moving averages help smooth out short-term data fluctuations and approximate seasonal fluctuations in disease incidence. Twelve-month moving averages can also be used to calculate another index of seasonal disease incidence known as specific seasonals. *Specific*

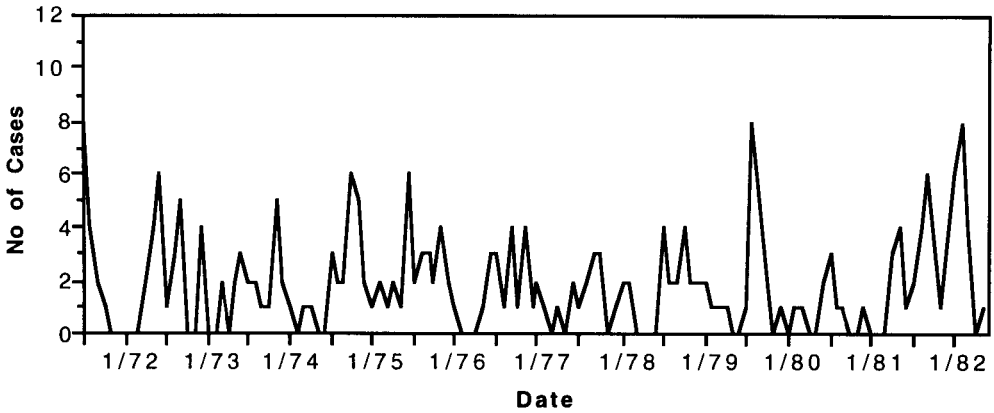


Figure 11.4 The occurrence and distribution of *Salmonella* cases among horses admitted to the Veterinary Medical Teaching Hospital, UC Davis, July 1971 to June 1982. (Reprinted with permission from Carter, J.D., Hird, D.W., Farver, T.B., and Hjerpe, C.A. 1986. Salmonellosis in hospitalized horses: seasonality and case fatality rates. *J.A.V.M.A.* 188:163-167.)

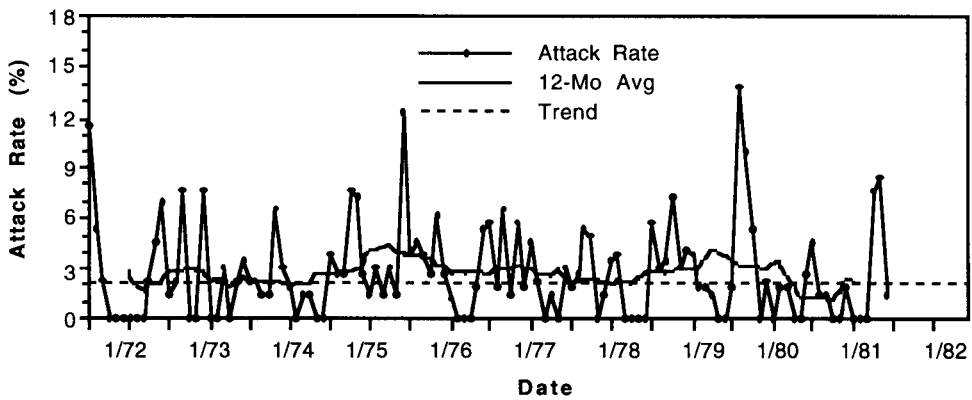


Figure 11.5 Monthly attack rate (incidence), 12-month centered moving average, and trend of salmonellosis in horses at the Veterinary Medical Teaching Hospital, UC Davis, July 1971 to June 1981. Monthly attack rate = (new cases) ÷ (daily average inpatients for the month). (Reprinted with permission from Carter, J.D., Hird, D.W., Farver, T.B., and Hjerpe, C.A. 1986. Salmonellosis in hospitalized horses: seasonality and case fatality rates. *J.A.V.M.A.* 188:163-167.)

seasonals are a ratio in which the observed monthly incidence rate is divided by the 12-month moving average incidence rate centered on the middle of that month (Schwabe et al, 1977). If specific seasonals are available for a number of years then they can be averaged (by mean or median) for each month to derive *typical seasonals*, which are indices of the amount of variation attributable to seasonal influences (Figure 11.6).

Subtraction of typical seasonals from specific seasonals leaves the combined cyclical and irregular variation in disease occurrence. *Cyclical changes* refer to the rise and fall of disease incidence with a periodicity of more than 1 year. *Irregular variation* reflects random or unpredictable variation in disease occurrence among individuals in a population. Both cyclical and irregular variation are associated with disease outbreaks.

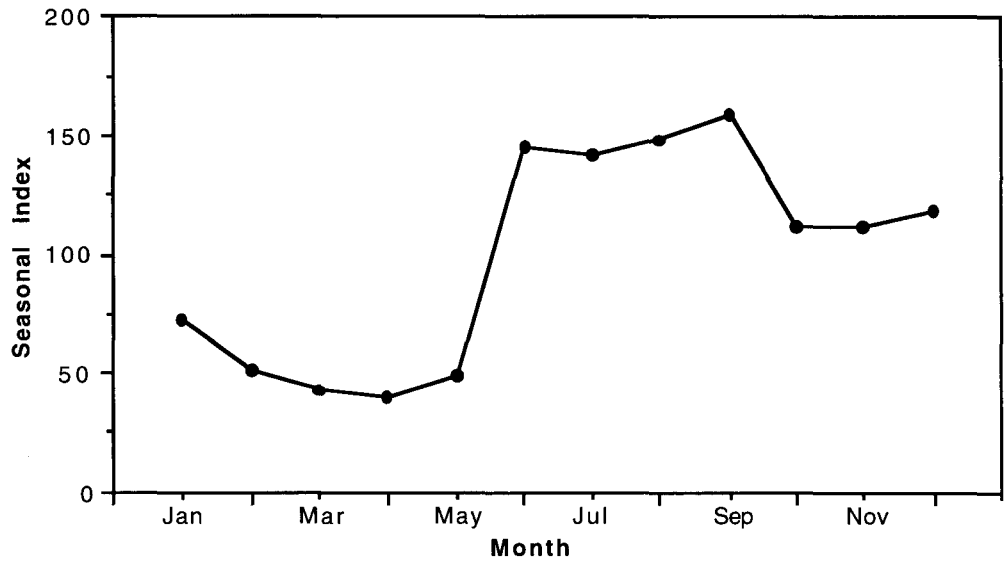


Figure 11.6 Seasonal index of *Salmonella* serotypes causing clinical disease in horses at the Veterinary Medical Teaching Hospital, UC Davis, July 1971 to June 1981. (Reprinted with permission from Carter, J.D., Hird, D.W., Farver, T.B., and Hjerpe, C.A. 1986. Salmonellosis in hospitalized horses: seasonality and case fatality rates. *J.A.V.M.A.* 188:163-167.)

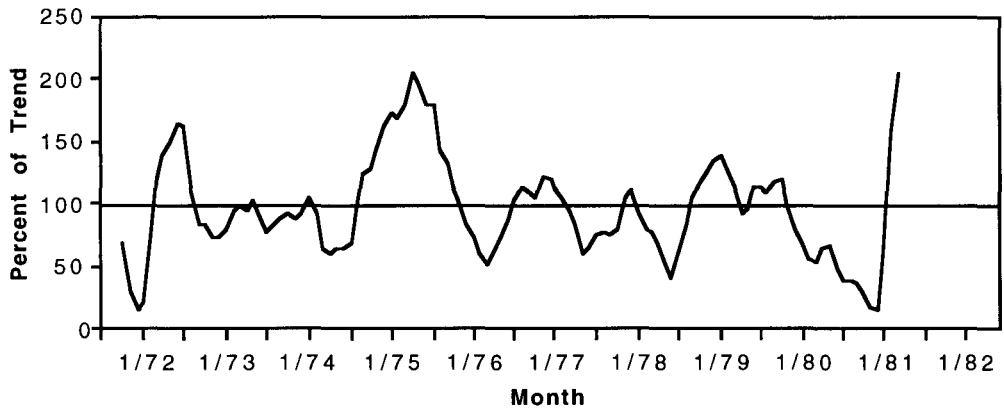


Figure 11.7 Cycles of salmonellosis in horses at the Veterinary Medical Teaching Hospital, UC Davis, July 1971 to June 1981. (Reprinted with permission from Carter, J.D., Hird, D.W., Farver, T.B., and Hjerpe, C.A. 1986. Salmonellosis in hospitalized horses: seasonality and case fatality rates. *J.A.V.M.A.* 188:163-167.)

Seasonal indices are expressed as percentage deviation from one. Thus, if the seasonal index were half the average for that month then it would be 50%; if it were twice the average it would be 200%. Examples of the application of time series analysis can be found in the following example.

EXAMPLE: In 1982 the entire Large Animal Clinic of the Veterinary Medical Teaching Hospital (VMTH) at the School of Veterinary Medicine, University of California at Davis,

was forced to close temporarily, because of a serious outbreak of *Salmonella saint-paul* infection in horses (Carter et al, 1986). An epidemiologic study of clinical salmonellosis during the 11-year period up to and including the outbreak (July 1971 through June 1982) revealed 245 cases of equine salmonellosis caused by 18 serotypes (Figure 11.4). The distribution of serotypes over time revealed disappearance of some serotypes and the introduction of others.

A time series analysis of monthly attack rates (number of new cases divided by daily average equine inpatient population for the month) revealed no significant overall increase or decrease in the rates (secular trend) over the 11-year period (Figure 11.5).

Seasonal fluctuations occurred, with highest incidence of salmonellosis from June through September, and lowest incidence from January through May (Figure 11.6). Cyclical changes appeared as three major outbreaks and several smaller outbreaks over the 11-year period (Figure 11.7). There was no regular pattern in the cycles that would be useful for forecasting salmonellosis outbreaks at the VMTH.

The result of the time series analysis of this outbreak may be summarized as follows: The incidence of salmonellosis in the VMTH has been stable over the past decade, neither increasing nor decreasing. There has been a definite seasonal trend with highest incidence from June through September and lowest incidence from January through May. Over the 10-year period from 1971 to 1981 there have been three major outbreaks and several smaller outbreaks. The contribution of any factors found to be associated with increased risk of salmonellosis should be interpreted in light of the temporal patterns of disease.

D. SPATIAL DISTRIBUTION

There are a number of ways of depicting 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. Another approach is the simple spot 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 of spatially distributed variables.

Figure 11.8 is an areal map that presents two types of information on the spatial distribution of Rocky Mountain spotted fever during 1985: its geographic range and the incidence per state (CDC, 1986a). When this information is compared with an areal map (Figure 11.9) of the spatial distribution of the two principal tick vectors in the United States, *Dermacentor andersoni* (the Rocky Mountain wood tick) and *D. variabilis* (the American dog tick), it is clear that *D. variabilis* plays the major role in human infections in the United States. Note that the dots in Figure 11.9 do not correspond to tick densities; only where each species has been reported and its probable range (Bishopp and Trembley, 1945).

IV. CASE STUDIES

A. CHARACTERISTICS OF VETERINARIANS IN ILLINOIS (SCHNURRENBERGER ET AL, 1972)

B. *BRUCELLA* INFECTIONS IN ILLINOIS VETERINARIANS (SCHNURRENBERGER ET AL, 1975)

A standard population is defined and used to adjust brucellosis prevalence rates among veterinary specialties.

1. Introduction

Most scientific reports pertaining to the health of veterinarians have been based on sero-

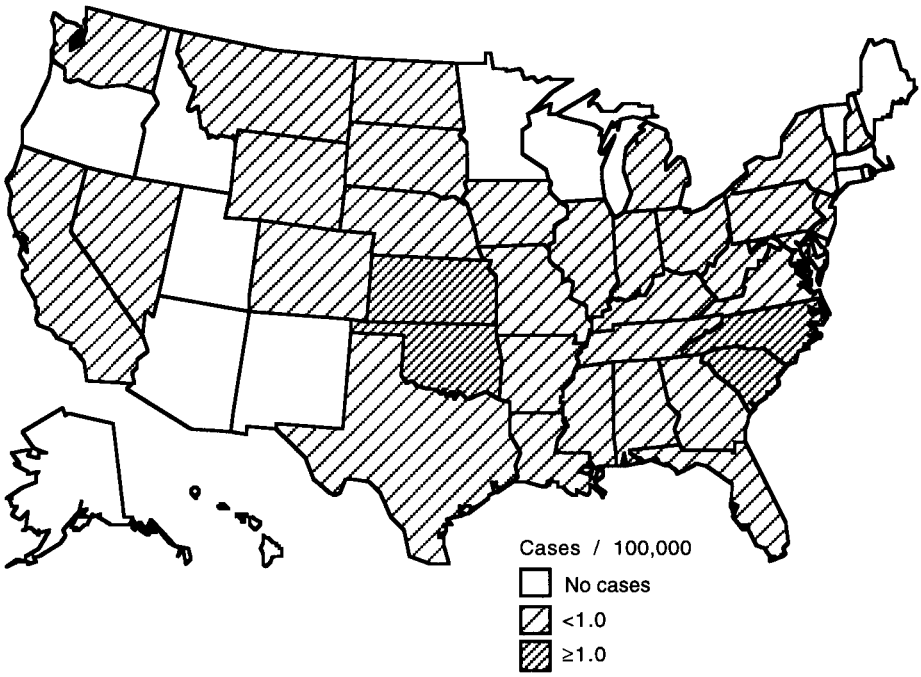


Figure 11.8 Reported Rocky Mountain spotted fever cases and rates by state, 1985. (Source: CDC. 1986a. Rocky Mountain spotted fever – United States, 1985. *MMWR.* 35 [Apr. 18, 1986; No. 15]:247-249.)

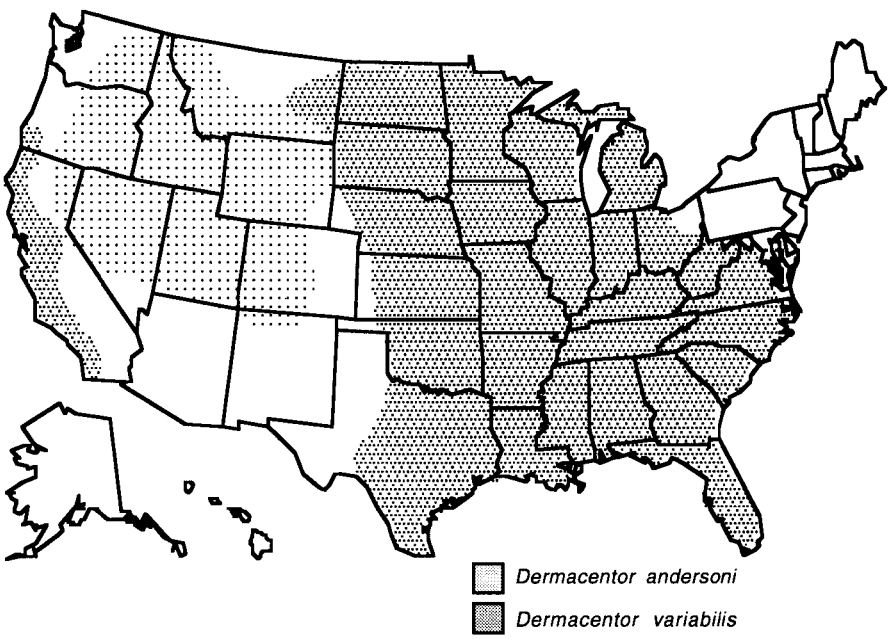


Figure 11.9 Distribution of *Dermacentor andersoni* and *Dermacentor variabilis* in the United States. (Reprinted with permission from Bishopp, F.C. and Trembley, H.L. 1945. Distribution and hosts of certain North American ticks. *J. Parasitol.* 31:1-54.)

logic studies of zoonotic diseases among veterinary students or private practitioners. Other studies have investigated rabies vaccinations, practice mobility, cause of death, economics and opinion leaders. Such studies are important to the future of the profession by pinpointing our problems, establishing trends and suggesting areas for improvement. A major problem with these studies has been the lack of base-line data on the entire profession, making it impossible to evaluate the representative nature of the populations studied.

2. Purpose of the Study

The first report (Schnurrenberger et al, 1972) is concerned with survey techniques for gathering such data and with certain characteristics of Illinois veterinarians. The follow-up paper uses these data to estimate age-adjusted prevalence rates for *Brucella* infections.

3. Epidemiologic Methodology

The initial population was assembled by cross matching all available lists of Illinois veterinarians from 1950 to 1967, using records of the AVMA, Illinois State Veterinary Medical Association (ISVMA), Illinois Department of Agriculture, Illinois Department of Registration and Education and the University of Illinois, College of Veterinary Medicine. Names were then removed from the initial population for any of the following reasons: death certificate on file with the Illinois Department of Public Health (IDPH), listed in the obituary columns of the *Journal of the American Veterinary Medical Association* or *Illinois Veterinarian*, known member of the armed forces while in Illinois, or known present address outside of Illinois.

Each of 1195 remaining veterinarians was surveyed through personal visits, telephone interviews or mailed questionnaires by county representatives of the ISVMA Auxiliary from July 1 to November 1, 1967. Each of the approximately 150 participating interviewers was provided with a 4-page instruction booklet describing the survey methodology. Each interviewer was requested to complete a questionnaire on every nonmilitary veterinarian in the area, regardless of whether a questionnaire had been preaddressed to the person. Questionnaires not completed by November 1 were returned to the IDPH where two trained clerical personnel conducted telephone interviews. The 1967 questionnaires were also completed by the 1968 graduating class of the College of Veterinary Medicine, University of Illinois.

Among the information requested in the 1967 survey were birthdate, practice type and history of illness with selected zoonoses. The survey was repeated in 1968 and 1969. In 1968 a detailed history of illness due to brucellosis was sought. The 1969 questionnaire included a question on zoonotic infections among family members. Information was transferred to IBM cards for processing. Each factor was cross-correlated with every other factor, and statistically significant associations noted.

4. Assumptions Inherent in the Methodology

The intent of the survey was to gather demographic data on every veterinarian permanently working in Illinois. It was assumed that the proportion of Illinois veterinarians not appearing on any of the lists would be relatively small, and that many would be picked up by interviewers at the county level. Validity of the brucellosis history of veterinarians depended on recall and assumptions about the sensitivity, specificity, concordance and duration of reactions to the brucellosis plate, card and tube tests.

5. Basic Epidemiologic Findings

A total of 1186 out of 1195 veterinarians responded to interviews in 1967, representing a cooperation rate of 99.2%. The age distribution of the standard population (all veterinarians, Figure 11.10) is compared with that for small animal practitioners, teachers and retirees (Figure 11.11).

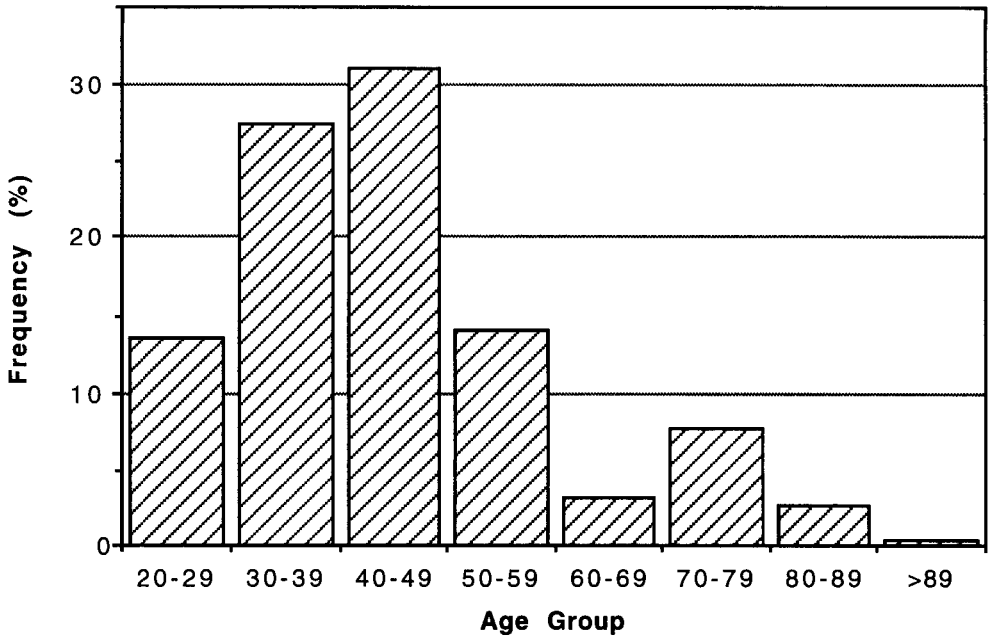


Figure 11.10 Age distribution for Illinois veterinarians, 1967. (Reprinted with permission from Schnurrenberger, P.R., Martin, R.J., and Walker, J.F. 1972. Characteristics of veterinarians in Illinois. *J.A.V.M.A.* 160:1512-1521.)

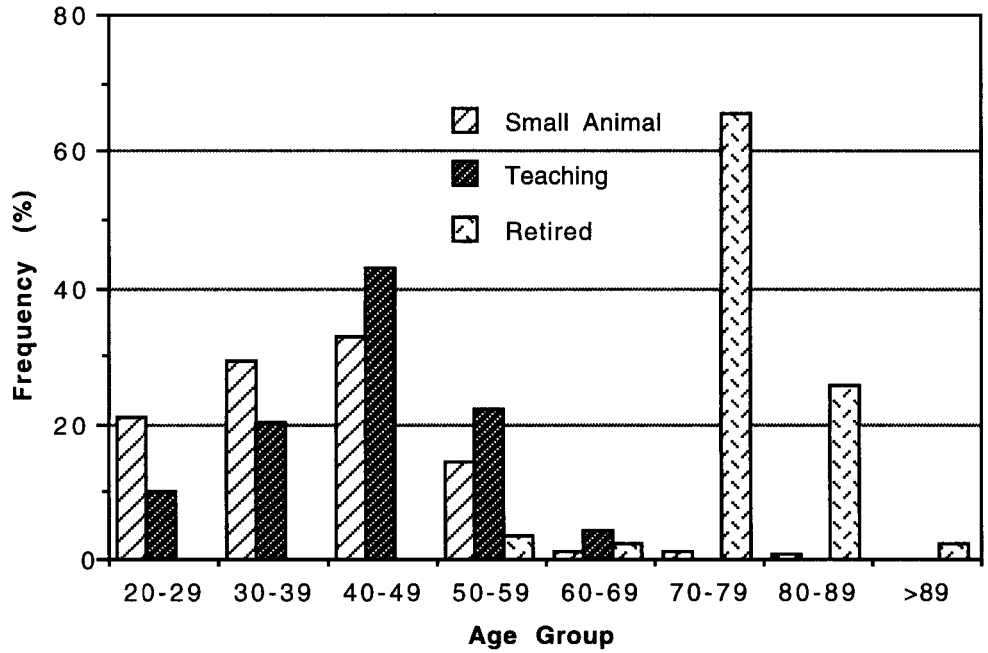


Figure 11.11 Age distribution for veterinary specialties. Illinois, 1967. (Reprinted with permission from Schnurrenberger, P.R., Martin, R.J., and Walker, J.F. 1972. Characteristics of veterinarians in Illinois. *J.A.V.M.A.* 160:1512-1521.)

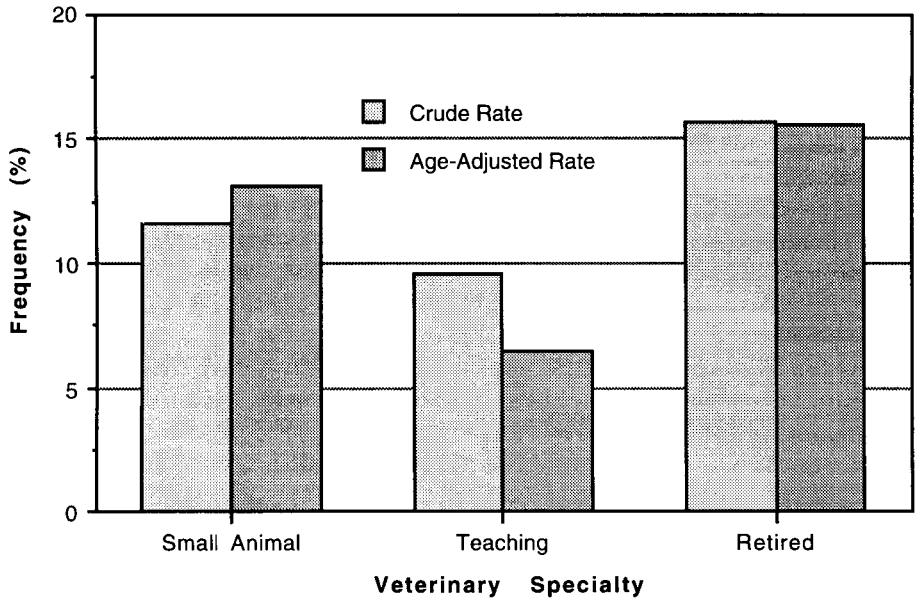


Figure 11.12 Prevalence of *Brucella* infections among Illinois veterinarians in selected specialties. (Reprinted with permission from Schnurrenberger, P.R., Walker, J.F., and Martin, R.J. 1975. *Brucella* infections in Illinois veterinarians. *J.A.V.M.A.* 167:1084-1088.)

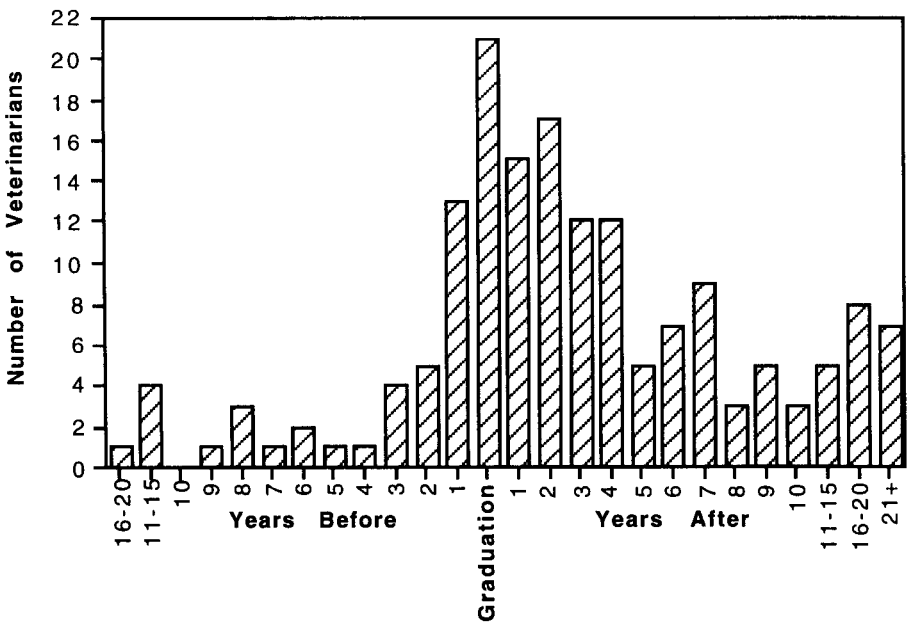


Figure 11.13 Relationship between onset of brucellosis and graduation in 156 Illinois veterinarians. (Reprinted with permission from Schnurrenberger, P.R., Walker, J.F., and Martin, R.J. 1975. *Brucella* infections in Illinois veterinarians. *J.A.V.M.A.* 167:1084-1088.)

The prevalence of brucellosis was subsequently compared among specialties based on the results of the 1967 questionnaire and a 16-year (1956, 1964, 1966, 1968, 1970, 1972) serologic survey of veterinarians (Schnurrenberger et al, 1975). Discrepancies between crude (age-unadjusted) and adjusted rates are depicted for three veterinary specialties (Figure 11.12) and illustrate the importance of using adjusted rates for comparison of groups. The age-adjusted prevalence of *Brucella* infection increased for small animal practitioners, decreased for teachers and remained the same for retirees, compared with crude rates. The prevalence of infection was found to be decreasing, as reflected by decreasing serologic reactor rates and by decreasing numbers of clinical diagnoses. Most of the small animal practitioners with a history of clinical brucellosis had been infected either as students or in an earlier practice type (Figure 11.13).

6. Conclusions and Measures Taken

The large number of untrained interviewers and the fact that many were related to veterinarians could have unfavorably influenced the validity of the answers given. However, close agreement between answers obtained by the auxiliary members and those obtained by trained interviewers suggests this bias was negligible. The cooperation rate of 99.2% reinforces the validity of the data in terms of its representative nature.

The difficulty in identifying and maintaining an accurate roster of veterinarians is undoubtedly due in great part to their mobility. In some cases, nonpractitioners feel that they do not belong to the profession, as illustrated by the occasional reply, "I used to be a veterinarian, but I'm not anymore."

The large animal and general practice groups appeared to be similar in the characteristics studied, whereas the other specialties frequently differed. This suggests that (1) the definitions used created artificial divisions, or (2) that similar factors determine whether a person enters and remains in these two practice types. Age seemed to be an important variable throughout the study. Differences in the age distribution of Illinois veterinarians were sufficiently great among most specialties to warrant comparisons after adjustment for age.

Strain 19 *Brucella* vaccine appeared to be increasing in relative importance as a source of infection for veterinarians. Part of the explanation for the decreasing infection rates following graduation might have been the existence of a group of veterinarians at high risk of infection because of personal habits ("klutzes"). Early infection, because of the high risk, would have resulted in rapid depletion of susceptible individuals from this group. As a result, in a few years the infection rate of the total veterinary population would no longer be dominated by this high-risk group, but would more nearly reflect the infection probability for the average veterinarian.

C. URBAN CATS: CHARACTERISTICS AND ESTIMATION OF MORTALITY DUE TO MOTOR VEHICLES (CHILDS AND ROSS, 1986)

Sources of bias in the estimation of geographic, temporal and host occurrence are addressed.

1. Introduction

Increases in the population of owned pets, estimated to number approximately 55 million cats and 52 million dogs in 1988 (Troutman, 1988), have brought concomitant increases in the number of free-ranging animals on city streets. These animals are at risk of death and trauma inflicted by motor vehicles. Some animal deaths on city streets may be due to causes other than vehicular trauma; however, the vast majority are automobile related.

Considerable animal suffering, as well as monetary expense, is associated with automobile-inflicted injuries. Free-ranging cats and dogs on city streets are the cause of many car accidents each year in the United States and abroad. Automobile-inflicted injuries are a source of additional monetary expense to municipal or private organizations, who must pick up dead or wounded animals.

2. Purpose of the Study

Although motor vehicles are believed to be a major cause of death or injury to free-ranging animals, accurate mortality data have not been gathered for cats or dogs, and few studies have described the characteristics of animals killed on city streets. In Baltimore, Municipal Animal Shelter (MAS) records are used to estimate yearly mortality of dogs and cats on city streets. This approach may seriously underestimate the number of animal deaths, because virtually all dead animals are picked up in response to telephone calls placed by city residents to the MAS, and unreported animals are not removed.

The purpose of this study by Childs and Ross (1986) was to obtain a better estimate of the number of cats dying on city streets by applying a mark-recapture technique, and to describe the significant characteristics of the population of cats dying on streets in Baltimore.

3. Epidemiologic Methodology

a. *Animals, Tagging, and Placement*

Geographic locations of dead cat pickups were plotted on a map of Baltimore, using a criss-cross directory of the city. From this map, the city was stratified by mortality density into areas of high (>4 dead cats), medium (3 or 4 dead cats) and low (0-2 dead cats) mortality for the mark-recapture study. The number of dead cats picked up by the MAS from 1980 to 1981 was used as a numerator for estimation of total mortality during the period of the mark-recapture experiment.

Adult stray cats that had been held at the MAS for the prescribed 5 days without adoption were euthanized, tagged and placed at the junction of two streets approximately in the center of selected census tracts between 11 PM and 3 AM. Fifteen dead cat placements, involving seven cats per placement (total = 105 dead cats), were made over a 1-year period. At each drop, three of the seven cats were placed in locations with low mortality designation, and two cats each were placed in individual locations with medium and high mortality designations.

b. *Monitoring of Dead Cat Pickups*

Animal shelter wardens were informed about the tagging project and were requested to note the date, location and tag number of each tagged cat retrieved. Wardens were not informed as to the actual date of individual dead cat placements.

c. *Characteristics of Street-Killed Cats*

Inspections were made from March 1980 through February 1981, about once a week (47 inspections), on a daily sample of dead cats removed from city streets by MAS wardens. Weight, sex, approximate age and sexual alteration or pregnancy in sufficiently intact cats were recorded. Cats weighing less than 2 kg were considered juveniles (<6 months of age).

d. *Statistical Methods*

Differences in distribution patterns of cat mortality based on city land use and season were analyzed by Chi-square tests. Total cats killed was estimated from the mark-recapture data using the formula

$$N = n(M/m)$$

where total population size (N) is estimated from the number sampled (n) and the number of marked animals (M) that are recaptured (m).

4. Assumptions Inherent in the Methodology

In the mark-recapture approach the authors assumed that (1) tags were permanent, (2) all tagged cats retrieved by the MAS were noted, (3) tagged dead cats were mixed randomly with

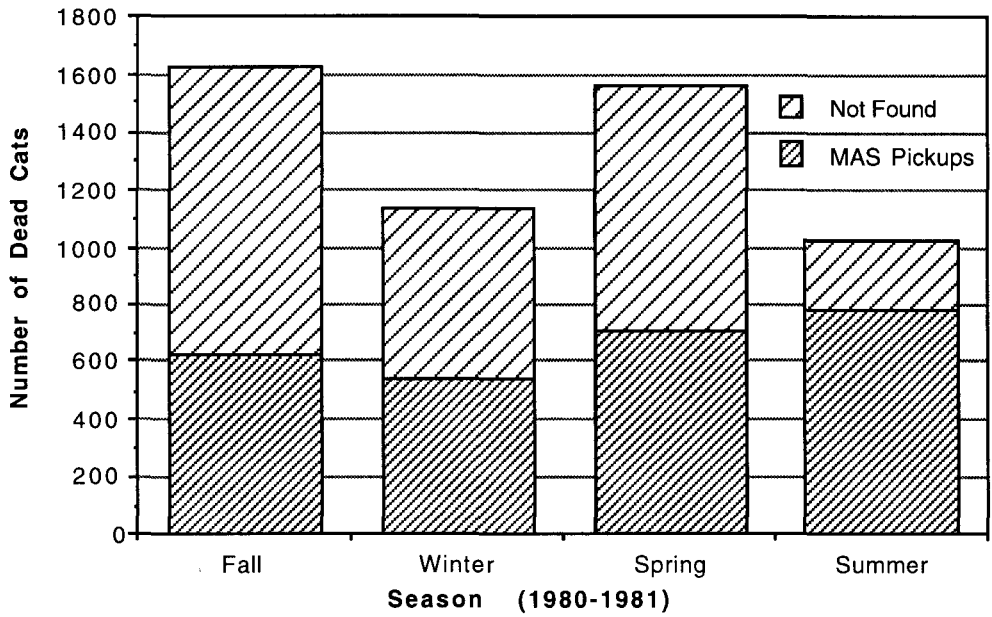


Figure 11.14 Raw and estimated numbers of dead cats occurring seasonally in Baltimore, MD. Raw numbers are represented as "MAS Pickups," which are seasonally increased by estimated numbers of cats that were "Not Found" (see Table 11.3) to yield a seasonally adjusted total number of dead cats. (Source of data: Childs, J.E. and Ross, L. 1986. Urban cats: characteristics and estimation of mortality due to motor vehicles. *Am. J. Vet. Res.* 47:1643-1648.)

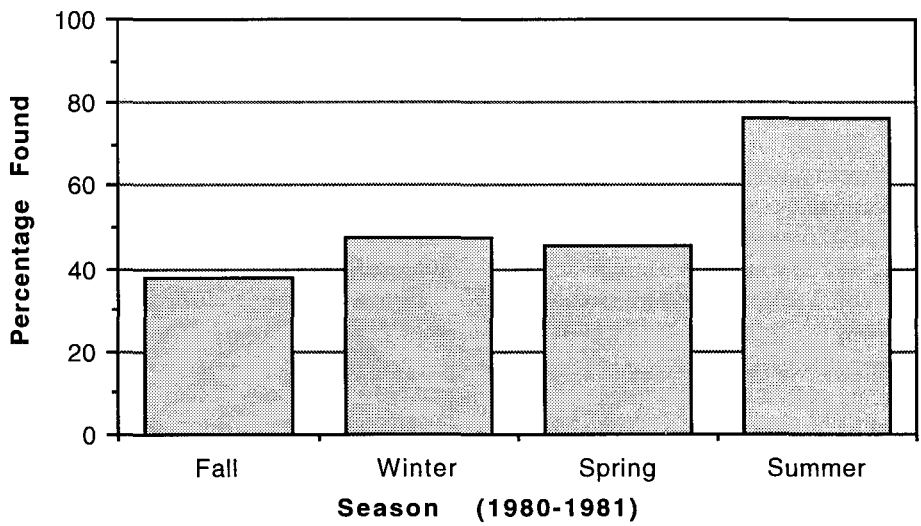


Figure 11.15 Seasonal pickup rate of dead cats. Mark-recapture experiment. (Source of data: Childs, J.E. and Ross, L. 1986. Urban cats: characteristics and estimation of mortality due to motor vehicles. *Am. J. Vet. Res.* 47:1643-1648.)

the natural population of dead cats and (4) tagged dead cats and naturally killed cats were equally likely to be retrieved by MAS crews or removed by other means.

Table 11.2 Estimated and actual seasonal and annual road-kill mortality figures for Baltimore city cats, based on tagged cat retrieval results and Municipal Animal Shelter records

Season-Year	*Number of Dead Cats Picked Up by MAS*	*Proportion of Tagged Dead Cats Picked Up (± SE)*	*Estimated Mortality for Seasonal Interval (95% CI)*
Fall 1980	620 ÷	0.3810 ± 0.1060 =	1627 (1052 – 3580)
Winter 1980-81	543 ÷	0.4762 ± 0.1090 =	1140 (787 – 2068)
Spring 1981	706 ÷	0.4524 ± 0.7680 =	1561 (1171 – 2339)
Summer 1981	780 ÷	0.7690 ± 0.9290 =	1024 (826 – 1345)
Year total (av)	2649 ÷	(0.5048 ± 0.0488) =	5248*

* Yearly total is estimated from proportion of tagged dead cats picked up and therefore is not identical to the sum of estimated mortality for the seasonal intervals.

Reprinted with permission from Childs, J.E. and Ross, L. 1986. Urban cats: characteristics and estimation of mortality due to motor vehicles. *Am. J. Vet. Res.* 47:1643-1648.

5. Basic Epidemiologic Findings

a. *Naturally Occurring Mortality*

The seasonal number of dead cat pickups by the MAS over a 1-year period is depicted in Figure 11.14. There was a consistent seasonality in the number of dead cats removed from city streets, with the majority being removed during the warm months of June through October. Overall, more dead cats were removed from residential areas than would be expected to occur by chance ($X^2 = 247.6$, $P < 0.001$, $df = 3$).

b. *Retrieval of Tagged Dead Cats*

Over a 1-year period, 53 (50.5%) of 105 tagged dead cats were retrieved by MAS wardens. An additional eight dead cats (7.6%) were reported by telephone to the MAS, but were no longer at the location when city wardens arrived. If cats picked up by and called in to the MAS were considered, then a minimal estimate of 58.1% (61/105) of the tagged dead cats were brought to the attention of city authorities. The remaining 44 dead cats were never reported to the MAS and were presumably removed or displaced by some other means.

Tagged dead cats were usually recovered by MAS wardens within a day of their placement on the street. There was no significant difference in the recovery rates of dead cats from areas of different mortality designations ($X^2 = 0.72$, $P = 0.70$, $df = 2$). However, the different seasonal rates of tagged dead cat pickups (Figure 11.15) approached statistical significance ($X^2 = 7.37$, $P = 0.06$, $df = 3$).

c. *Estimation of Cat Mortality*

The total number of dead cats picked up by the MAS during the time of the tagging experiment was 2649. The adjusted frequency of cat deaths is estimated in Table 11.2 and depicted in Figure 11.14, each bar representing the sum of the reported ("MAS Pickups") and unreported ("Not Found") cats for each season. The largest estimated cat mortality occurred in fall

and spring (1627 and 1561 dead cats, respectively), in contrast to the pattern derived solely from analysis of MAS records. The overall estimated annual street mortality for cats in Baltimore was 5248, approximately twice the number actually removed by MAS personnel.

d. *Characteristics of Dead Cats*

A significantly larger proportion of the 212 dead cats sampled over the 1-year study period was male (63.2%, $P < 0.05$, based on the assumption of a 1:1 sex ratio). At least 20.3% were presumed to have been owned at one time, based on sexual alteration or presence of a collar. Juvenile or young adult animals made up only 18.4% of the entire dead-animal sample, based on body weight less than 2.0 kg and tooth eruption patterns. Necropsy revealed that 37% of females were reproductively active, based on pregnancy or lactation. The vast majority of the dead cats examined were domestic short hairs (98.1%), and only two domestic long hairs and two Siamese were found in the sample.

6. Conclusions and Measures Taken

The investigators were able to assess the role of reporting and sampling bias in the estimation of geographic and temporal occurrence of feline mortality through the use of a mark-recapture technique. Data from the mark-recapture experiment indicated that the summer peak in feline mortality, based solely on MAS records, was an artifact attributable to the greater percentage of dead cats reported and picked up in the summer. Urban residents may rely on MAS crews during warm months when dead cats rapidly decompose, but may independently dispose of cat corpses during cooler months. The authors estimate that over 5248 free-ranging cats are killed by automobiles each year in Baltimore, but that strong evidence for a seasonal trend does not exist.

The geographic distribution of street-killed cats is significantly associated with the residential areas within Baltimore. Higher feline mortality was associated with high human population densities and is most likely related to increased absolute numbers of owned animals. Reporting bias apparently did not influence this finding, as dead cats were reported at similar rates, regardless of their placement in the city.

Earlier census data from 440 households in two locations in Baltimore indicated that 13.4% of households own an average of 1.24 cats per cat-owning household. Extrapolating these data to the estimated 281,414 housing units in Baltimore in 1980, the authors estimated that there are approximately 46,759 owned cats in the city. Furthermore, it was previously reported that 40.5% of owned cats were allowed to range freely on city streets, and that 64.7% of these would be recognizable as owned cats on the basis of sexual alteration and/or presence of a collar. If the 20.3% estimate for previously or presently owned cats among the street kills is adjusted in light of the census survey, then approximately 31.4% of street kills would be previously or presently owned cats. It follows that 1648 of the 5248 estimated dead cats removed annually from city streets were owned at one time. Thus, the annual incidence of street mortality in the total owned cat population is on the order of 1648/46,759 (3.5%), and the annual incidence of street death in owned cats allowed to range freely is 1648/46,759 x 0.405, or approximately 8.7%. These are crude estimates and are generated only to indicate the possible order of magnitude of the problem.

These data provide a more realistic estimate of street deaths than can be obtained by direct examination of MAS records and provide insight into the population of free-ranging animals in Baltimore that are at risk of death and trauma due to motor vehicles.

V. SUMMARY

Occurrence refers to the frequency distribution of disease over space (spatial or geographic occurrence), time (temporal occurrence) or within a host population. This information is use-

ful not only to gain a better appreciation of the significance of the disease, but may suggest the probable cause, source and mode of transmission of the condition. The first step in any disease investigation is identification of the cases and noncases. Cases may be defined on the basis of a discrete set of signs and symptoms, performance indicators or epidemiologic criteria. Epidemiologic criteria, such as the occurrence of the disease, may be added to the case definition.

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. An attack rate measures the proportion of the population that develops disease among the total exposed at the beginning of the outbreak. The attack rate is essentially an incidence rate where the time period of interest is the duration of the epidemic.

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 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.

Most diseases have characteristic patterns of temporal occurrence. A disease is sporadic when it occurs rarely and without regularity in a population unit. 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, as reflected in terms used to describe the levels of occurrence of endemic disease: (1) holoendemic, when most animals are affected, (2) hyperendemic, when a high proportion of animals are affected, (3) mesoendemic, when a moderate proportion of animals are affected or (4) hypoendemic, when a relatively small proportion of animals are affected. A disease is epidemic when its frequency within the population during a given time interval is clearly in excess of its expected frequency.

Time series analysis is concerned with the detection, description and measurement of patterns or periodicities from temporal occurrence data. 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 (incidence) are influenced by one or more of the following: (1) secular trend, (2) seasonal fluctuation, (3) cyclic variation and (4) irregular variation. Secular trends are overall long-term rises or declines in incidence rate that occur gradually over long periods of time. Seasonal fluctuations are regular changes in incidence rates with periods shorter than a year and may be expressed as specific or typical seasonals. Subtraction of typical seasonals from specific seasonals leaves the combined cyclical and irregular variation in disease occurrence. Cyclical changes refer to the rise and fall of disease incidence with a periodicity of more than 1 year. Irregular variation reflects random or unpredictable variation in disease occurrence among individuals in a population. Both cyclical and irregular variation are associated with disease outbreaks.

There are a number of ways of depicting the spatial distribution of disease frequency. Areal maps depict the distribution and frequency of disease within defined boundaries, as counties, states or ecological zones. Another approach is the simple spot map, where each dot either represents a case or is scaled to represent the frequency of disease.

Chapter 12

ESTABLISHING CAUSE

I. 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.

II. MULTIPLE CAUSATION OF DISEASE

Determinants of disease include both the etiologic agents 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.

A. AGENT FACTORS

The biological properties of agents, such as pathogenicity and virulence, strains and genetic drift, are primary determinants of the ability of an agent to cause disease. Contributors to the pathogenicity and virulence of disease agents are generally covered in microbiology texts and are not discussed here.

B. 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 shortly after being moved to a new location.

Determinants of disease include the agent, host and environment.

Populations also differ in susceptibility. Resistance in populations is called *herd 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 previous

Table 12.1 Relationship between a pathogen's intrinsic reproductive rate (R_0) 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_0	*P (%)*
Smallpox	Developing countries before global eradication campaign	3.5	70-80
Measles	England and Wales (1956-68) Parts of United States (1910-30)	13	92
Whooping cough	England and Wales (1942-50) Maryland (1908-17)	17 13	94 92
German measles	England and Wales (1979) West Germany (1972)	6	83
Chicken pox	Parts of United States (1913-21; 1943)	9-10	90
Diphtheria	Parts of United States (1910-47)	4-6	80
Scarlet fever	Parts of United States (1910-20)	5-7	80
Mumps	Parts of United States (1912-16; 1943)	4-7	80
Poliomyelitis	Holland (1960); United States (1955)	6	83
Malaria (*Plasmodium falciparum*)	Northern Nigeria (1970s)	80	99
Malaria (*Plasmodium malariae*)	Northern Nigeria (1970s)	16	94

R_0 = 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., $R_0 (1 - P) < 1$.

Reprinted by permission of American Scientist, journal of Sigma Xi, The Scientific Research Society from May, R.M. 1983. Parasitic infections as regulators of animal populations. *Am. Scientist.* 71:36-45.

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.

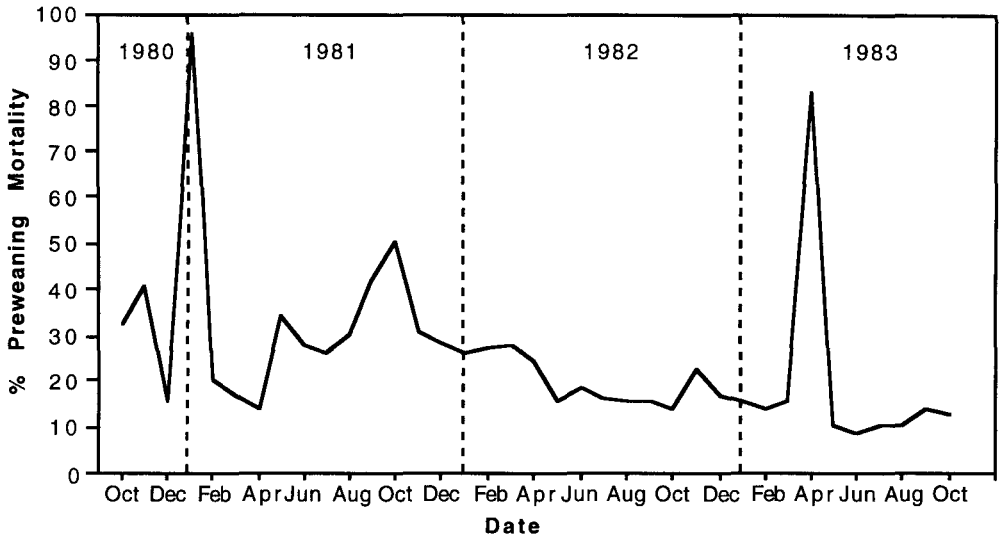


Figure 12.1 Percent preweaning mortality over a 3-year period in a 130- to 220-sow farrow-to-finish herd. (Reprinted with permission from Morrison, R.B. and Joo, H.S. 1985. Prenatal and preweaning deaths caused by pseudorabies virus and porcine parvovirus in a swine herd. *J.A.V.M.A.* 187:481-483.)

Increased herd immunity has the effect of limiting the spread of directly transmitted diseases by reducing the proportion of *effective contacts*, e.g., contacts between infected and susceptible animals that result in transmission of a disease agent. Increased herd immunity may also limit the spread of indirectly transmitted and airborne disease agents by reducing environmental contamination. In either case, the reproductive rate for a disease agent may fall below that required for its maintenance in the population, leading to its eventual eradication.

It follows that the higher the *intrinsic reproductive rate* (R_0; see Chapter 13) of a disease organism, the higher the level of herd immunity that must be achieved for its eradication. Very high levels of artificially induced herd immunity are required to eradicate diseases whose intrinsic reproduction rates are high (Table 12.1). The relatively small value of R_0 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_0 for malaria 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.

EXAMPLE: Pseudorabies is a viral infection caused by *Herpesvirus suis*. Infection occurs naturally in virtually all mammals except humans and certain primates. The disease in swine may cause abortion among pregnant sows and increased mortality among baby pigs and weanlings. The principal mode of transmission is direct contact between infected and susceptible swine.

Sequential outbreaks of pseudorabies virus, manifested primarily as preweaning mortality, were documented over a 3-year period at a swine farm in southern Minnesota (Figure 12.1). The herd was housed at a farrow-to-finish facility. The females were housed in four groups according to their stage of pregnancy. Each group was penned outside on concrete, with shelter provided. Wire fence lines allowed nose-to-nose contact between groups. The breeding period

for each group was restricted so that all-in/all-out farrowing, nursery and grower-barn schedules could be maintained.

Major outbreaks of preweaning mortality occurred in January 1981 and April 1983. A serologic survey conducted in March 1983 of 40, 4- to 6-month-old hogs in the finishing barn and 10 replacement gilts failed to reveal antibody to pseudorabies virus. The all-in/all-out movement of pigs from farrowing to finishing may have inhibited the spread of pseudorabies virus, thereby reducing herd immunity to pseudorabies infection. Pseudorabies virus-susceptible gilts gradually replaced immune sows, which further reduced herd immunity (Morrison and Joo, 1985).

C. 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 quantitate and manipulate. Examples are the influence of milking hygiene on the occurrence of bovine mastitis and management practices on neonatal calf mortality (see Table 12.3).

III. MULTIPLE CAUSATION AND KOCH'S POSTULATES

In 1882 Koch set forth the following postulates for determining that an infectious agent is the cause of a disease (Fletcher et al, 1982):

- The organism must be present in every case of the disease.
- The organism must be isolated and grown in pure culture.
- The organism must, when inoculated into a susceptible animal, cause the specific disease.
- 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.

The causes of many diseases cannot be established by means of Koch's postulates.

EXAMPLE: Enzootic pneumonia of calves is an infectious respiratory disease of calves maintained in confinement, either indoors or outdoors. Morbidity rates may approach 100% and mortality rates frequently exceed 20%. 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 virus followed by (3) a superinfection with any of a variety of bacteria. 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 (Fletcher et al, 1982). Fortunately, other criteria may be applied to test the strength of a presumed cause-effect relationship.

IV. ESTABLISHING CAUSE

A. STRENGTH OF STUDY DESIGNS

In Chapter 1 a variety of epidemiologic study designs were described. Generally, as one goes down the list in Table 1.4 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.

B. 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 proved useful in establishing cause in clinical practice and outbreak investigation.

EXAMPLE: The temporal relationship is intrinsic to the definition of nosocomial infections. Nosocomial infections have been defined as those infections that are produced by microorganisms acquired during hospitalization. Infections incubating at the time of the patient's admission to the hospital are not considered nosocomial (Koterba et al, 1986). An awareness has developed in the veterinary profession regarding the importance of nosocomial infections with bacteria resistant to antimicrobials.

During the spring of 1983, it became apparent that a number of neonatal foals were developing serious infections while hospitalized in the VMTH of the University of Florida. The increased incidence of disease was initially ascribed to prematurity and failure of passive antibody transfer. However, after evaluation of the antibiotic susceptibility patterns of bacteria isolated from the foals and the antibiotic regimens used, it was thought possible that these foals were being infected by multidrug-resistant gram-negative bacteria (lactose fermenting Enterobacteriaceae) residing within the large animal hospital. Consequently two studies were conducted: (1) a baseline study of 6 months duration to determine the prevalence of multiresistant bacteria and incidence of nosocomial infections in the equine hospital, and (2) a prospective study of 3 1/2 months duration of the changes in antibiotic resistance patterns of bacteria isolated from the feces of horses after 7 days of hospitalization. Day seven was chosen for collection of the second specimen to allow sufficient time for evacuation of all intestinal tract contents present at the time of admission. Isolates of *Salmonella* were eliminated from both studies because horses suspected of *Salmonella* infection were routinely kept in isolation.

An infection was considered community acquired if it was isolated from specimens obtained from the horse on the day of admission to the hospital. An infection was considered nosocomial if the specimen for bacterial culturing was taken on day two or later of the horse's hospital stay and fit any one of the following four situations:

(1) Bacterial culture of specimens taken at the time of admission yielded no microbial growth or grew bacteria of a different type or with a different susceptibility pattern than those taken after day two.
(2) The bacteria were isolated from an indwelling intravenous catheter site.
(3) A mare or foal healthy on admission and accompanying a sick mare or foal subsequently developed an infection that was proven via bacterial culturing.
(4) The bacteria were isolated from a surgical wound.

Infections were considered not identified if

(1) No specimens were obtained early for bacterial culturing.
(2) The medical records were insufficient for proper evaluation of the case.
(3) The horse had been previously hospitalized in the hospital facility.

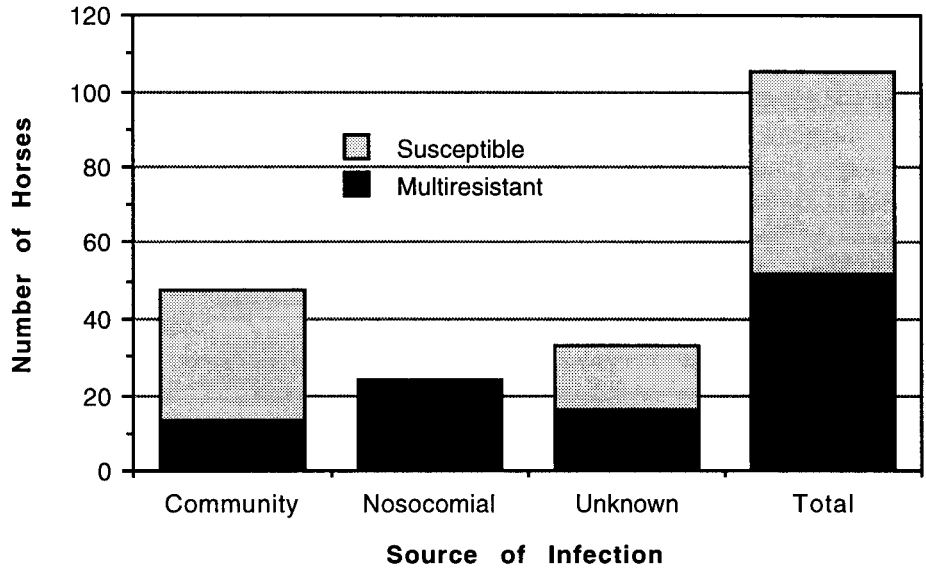


Figure 12.2 Results of a cross-sectional survey of gram-negative aerobic bacteria (excluding *Salmonella*) isolated from horses based on susceptibility to antibiotics and presumed source of infection. (Source of data: Koterba, A., Torchia, J., Silverthorne, C., Ramphal, R., Merritt, A.M., and Manucy, J. 1986. Nosocomial infections and bacterial antibiotic resistance in a university equine hospital. *J.A.V.M.A.* 189:185-191.)

Antimicrobial susceptibility tests were done by the Kirby-Bauer disk diffusion method. Any bacterial isolate resistant to at least five antibiotics used in routine susceptibility tests for gram-negative bacteria at the VMTH was considered multiresistant.

For the prospective study, the Wilcoxan signed rank test was used to compare the number of antibiotics to which a bacterial species was resistant on day 1 versus day 7. This test is appropriate since the authors are not asking whether there is a difference in the number of isolates resistant to antibiotics (which could be tested with a Chi-square test), but rather whether there is a difference in the number of antibiotics that isolates are resistant to. A patient may have the same microorganisms isolated on day 7 as day 1, but the day 7 isolates may be resistant to a greater number of antibiotics.

To determine whether the nosocomial infections were clustered in a certain part of the VMTH, a diagram was made of the stalls that housed the infected horses. To determine possible sources of infection, specimens were taken from water, handwashing soap, doormats, stocks, air, recovery stall and the hands of clinicians and students for bacterial isolation.

Specimens from 109 of 677 horses admitted over the 6-month study period contained gram-negative aerobic bacteria other than *Salmonella*. Antibiotic susceptibility tests were performed on 105 of the 109 horses. Twenty-three (21.9%) of the 105 horses had developed nosocomial gram-negative aerobic infections other than *Salmonella*, with high rates of resistance to gentamicin, kanamycin and trimethoprim-sulfadiazine, three of the most frequently prescribed antibiotics in the hospital. Multiresistant bacteria were implicated in 96% of nosocomial infections compared with 27% of community acquired infections and 48% of infections of unknown origin. Multiresistant bacteria comprised 50% of all gram-negative aerobic isolates (Figure 12.2).

Escherichia coli and *Klebsiella* spp. were the only Enterobacteriaceae found in sufficient numbers in the prospective study for comparison and statistical analysis (Table 12.2). Of 24

Table 12.2 Prospective study of changes in resistance patterns of bacterial intestinal flora of horses during 7 days of hospitalization

Bacteria	*Treated with Antibiotics*	*No. of Isolates*	*Mean No. of Antibiotics (with Range) to Which Bacteria Were Resistant*		*P Value**
			Day 1	*Day 7*	
E. coli	Yes	16	2.06 (1-6)	3.94 (1-9)	<0.014
	No	8	2.12 (0-6)	3.5 (1-7)	NS
			Mean = 2.08 (0-6)	Mean = 3.83 (1-9)	<0.003
Klebsiella sp.	Yes	2	2.5 (1-4)	6 (5-7)	NS
	No	5	3.8 (3-5)	4.6 (4-5)	NS
			Mean = 3.43 (1-5)	Mean = 5 (4-7)	<0.043

*Wilcoxon signed rank test of change in number of antibiotics bacteria were resistant to on day 7 versus day 1. NS denotes not significant or not performed because of sample size.

Reprinted with permission from Koterba, A., Torchia, J., Silverthorne, C., Ramphal, R., Merritt, A.M., and Manucy, J. 1986. Nosocomial infections and bacterial antibiotic resistance in a university equine hospital. *J.A.V.M.A.* 189:185-191.

paired fecal specimens obtained, isolates of both species of bacteria on day 7 were resistant to a significantly higher number of antibiotics than day 1 isolates ($p = 0.003$, $p = 0.043$, respectively).

The authors concluded that the incidence of nosocomial infections was not unacceptably high, but that a serious problem with highly resistant bacterial infections did exist in the equine hospital. The prospective paired fecal bacterial study supported the hypothesis that while hospitalized, the patients' intestinal tracts were being colonized by a more resistant group of gram-negative bacteria. The most important sequela was considered to be prolongation of hospital stay.

C. 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. Earlier chapters have shown how the strength of association can be measured by comparing relative risk, odds ratios or correlation coefficients. Another way to evaluate the strength of an association is analysis of variance. This statistical technique permits one to compare mean values for more than two groups while adjusting for variation within each group. The following is an example of using statistical techniques to explore multiple causes of disease based on strength of association.

EXAMPLE: Neonatal calf mortality represents a major economic loss to the cattle industry, with estimates of mortality ranging from 8% to 25%. A questionnaire survey of 477 Michigan dairy herd operators attending extension meetings was carried out to determine what factors influenced dairy calf mortality (Oxender et al, 1973).

Three categories of calf mortality were identified: (1) dead or died during birth, (2) living at birth but dead before 14 days of age and (3) died between 15 and 60 days of age. All rates were based on the total number of calves born and included both live and dead births. A one-way analysis of variance with unequal numbers was used to analyze the results. Scheffe's test was

Table 12.3 Factors evaluated for their association* with dairy calf mortality in Michigan

Factor	No. of herds	Births (%)		Deaths (%)		(%) Total Mortality
		Live	Dead	0-14 days	15-60 days	
Herd size						
<50	217	93.9	6.1	7.5	2.5	16.1a
50-100	199	93.6	6.4	8.8	2.9	18.1a,b
100-200	56	92.5	7.5	10.6	2.8	21.1a,b
>200	5	89.6	10.5	18.1	6.3	34.9b
All	477	93.6	6.4	8.5	2.8	17.7
First feeding of colostrum						
<6 hours	267	N/A	N/A	7.6c	2.6	10.2c
6-12 hours	151	N/A	N/A	10.5d	2.9	13.4d
Days colostrum fed						
0	6	N/A	N/A	19.7e	2.4	22.1
1	22	N/A	N/A	8.4e,f	2.7	11.1
2	89	N/A	N/A	10.9e,f	3.2	14.1
3	345	N/A	N/A	7.8f	2.7	10.5
Type of housing						
Stanchion	125	94.1	5.9g	6.7	2.1	14.7g
Free stall	259	93.5	6.5h	9.6	3.2	19.3h
Loose housing	31	93.0	7.0g,h	10.3	3.5	20.9g,h
Use of antibiotics						
No	236	94.2	5.9i	7.2i	2.2i	15.3i
Yes	227	92.9	7.1j	9.8j	3.5j	20.4j

Other significant factors ($P < 0.05$)

Calves raised in area separate from maternity stalls	Reduced 4.6%

Insignificant factors ($P > 0.05$)

Feeding of whole milk rather than milk replacer	Reduced 2.1%
Pails rather than bottles to feed calves	No change
Use of thermometer to take temperature of sick calves	Reduced 1.6%
Person responsible for feeding neonatal calves	
Hired man	19.9%
Son or daughter	18.2%
Owner	18.0%
Wife	16.6%
Jersey herds	20.9%
Guernsey herds	19.4%
Holstein-Friesian herds	17.7%

*Different letters indicate significant ($P < 0.05$) differences in column.

Reprinted with permission from Oxender, W.D., Newman, L.E., and Morrow, D.A. 1973. Factors influencing dairy calf mortality in Michigan. *J.A.V.M.A.* 162:458-460

used to test columns for significant differences when more than two factors were being compared.

The mean number of calves born per herd was 70.1. Total mortality (attack rate) for 477 herds was 17.7%, broken down as 6.4% at birth, 8.5% at 0 to 14 days and 2.8% at 15 to 60 days. Diarrhea (70% of the herds) and pneumonia (41% of the herds) were the principal causes of calf deaths.

A list of factors analyzed for the strength of their association with calf mortality appears in Table 12.3. Herd size, use of colostrum during the postnatal period, separation of calf-rearing areas from maternity stalls and antibiotic usage were significantly associated with calf mortality. Herds housed in free stalls had significantly greater calf mortality than those housed in stanchion barns. However, type of housing and herd size were closely related, and were thus confounding variables. Antibiotic usage was significantly associated with increased calf mortality. The most likely interpretation is that antibiotics were used in an attempt to solve problems rather than prevent them.

Inadequate planning, overcrowding, lack of colostrum feeding, poor ventilation and labor shortages seemed to be the most common problems associated with the larger herds. Increases in population density of animals in the larger herds may also have contributed to spread of bacterial and viral infections.

D. 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*. 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.

EXAMPLE: The feeding of low concentrations of certain antibiotics improves feed efficiencies and increases the rate of gain in swine production. However, feeding of antibiotics has also been implicated in an increase in resistance to antibiotics of organisms such as *Salmonella*. Identification of *Salmonella* serotypes and determination of their antimicrobial susceptibility pattern are necessary for implementation of an efficacious treatment schedule.

From 1979 through 1983, 277 *Salmonella* isolates (27 serotypes) were recovered from swine necropsied at Kansas State University (Mills and Kelly, 1986). *Salmonella choleraesuis* was the most common isolate, making up 66.4% of the total. The resistance patterns of *Salmonella* to most antimicrobial agents did not change from 1979 to 1983. Carbadox was a notable exception, as indicated by a progressive annual decrease in percentage of isolates that were susceptible (Figure 12.3). Carbadox has been used as a feed additive since 1972 to prevent dysentery (due to *Treponema hyodysenteriae*), but is approved for use in swine with *Salmonella*. The authors concluded that the percentage of *Salmonella* isolates resistant to Carbadox appeared to be increasing (Mills and Kelly, 1986). The hypothesized cause-effect relationship is that prolonged exposure to the antimicrobial (the "dose") causes emergence of increasing numbers of resistant bacteria through selection (the "response").

E. BIOLOGICAL PLAUSIBILITY

Epidemiologic study designs are particularly 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

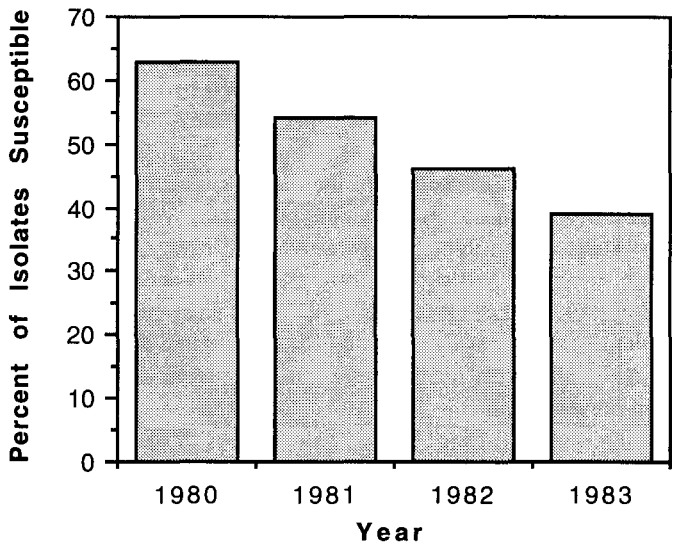


Figure 12.3. Stepwise decrease in *Salmonella* isolates susceptible to Carbadox from 1980 to 1983. (Reprinted with permission from Mills, K.W. and Kelly, B.L. 1986. Antibiotic susceptibilities of swine to *Salmonella* isolates from 1979 to 1983. *Am. J. Vet. Res.* 47:2349-2350.)

explanation does not necessarily mean that a causal association is absent. It may simply mean that current medical knowledge is incomplete.

F. CONSISTENCY

Evidence for a causal relationship is strengthened when several studies conducted under different conditions all come to the same conclusion. An example can be found in Chapter 8, where clinicians from several parts of the country and in distinct practice settings concurred on the beneficial effects of a new analgesic for equine colic (Gingerich et al, 1985). On the other hand, inconsistency in clinical findings may sometimes be attributed to differences in study design.

G. 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.

EXAMPLE: On July 24, 1982, a disease outbreak affected 43 of 67 lactating dairy cows over an 11-day period (Abbitt et al, 1984). Signs of disease included weakness, ataxia, drooling, inability to rise ("downer cows") and death without agonal movements or respirations. Many of the dead cows were in sternal recumbency, with the head bent to the flank or maintained forward. Necropsy findings included mucoid enteritis and patchy areas of congestion/hyperemia in the small intestinal mucosa. The latter sometimes produced a dramatic "tiger stripe" pattern.

A team of veterinarians from the Texas Veterinary Medical Diagnostic Laboratory (TVMDL) visited the farm on the fifth day of the outbreak. The herd was divided clinically (clinical staging) into three groups: group 1 (six clinically normal, lactating cows), group 2

Table 12.4 Rule-out list of possible causes for catastrophic death losses in a dairy herd

Suspected Etiologic Agent	*Hypothesized Exposure*	*Test Results and Comments*
Nitrate intoxication	Sorghum pasture	Negative
Prussic acid intoxication	Sorghum pasture	Negative
Hypocalcemia	Milk production	Negative
Hypomagnesemia	Nutritional deficiency	Negative
Arsenic intoxication	Environmental	Negative
Lead intoxication	Environmental	Negative
Aflatoxins	Feed	Negative
Urea	Feed	Negative
Insecticides	Feed, environment	Negative
Rat bait (warfarin and sulfanilamide)	Feed, water	Not accessible to cattle
C. botulinum toxin in rancid bone meal or feed	Feed	Did not produce disease when fed to laboratory animals
C. botulinum toxin in stagnant runoff water from milking parlor	Environment	Domestic ducks and horses unaffected

Source of data: Abbitt, B., Murphy, M.J., Ray, A.C., Reagor, J.C., Eugster, A.K., Gayle, L.G., Whitford, H.W., Sutherland, R.J., Fiske, R.A., and Pusok, J. 1984. Catastrophic death losses in a dairy herd attributed to type D botulism. *J.A.V.M.A.* 185:798-801.

(seven lactating cows identified as incipient cases by the owner) and group 3 (six downer cows). Samples of blood, feces and urine were obtained from cows in each of the three groups. Samples of grain, forage, bone meal and water available to the lactating cows were also obtained for analysis.

Cows in group 1 appeared normal. Group 2 cows appeared to have reduced fill of the gastrointestinal tract and were more resistant to urination induced by "feathering." Group 3 cows were alert and most maintained themselves in sternal recumbency. The owner reported that they had no obvious difficulty eating or drinking. Rectal temperature, heart rate and respiratory rates were normal, but some of the cows were "clammy" to the touch. Withdrawal reflexes of the hindlimbs (tested by pinpricks) were weak or absent. Pupillary response to light seemed slow, and rumen motility was absent or extremely weak. Peristaltic waves were evident by rectal palpation, and small piles of feces coated by a thick layer of gelatinous mucus were behind each cow. Group 3 cows would not urinate in response to feathering, although several had markedly distended bladders. A month after the outbreak some of the surviving cows were still weak and ataxic.

The cause of this catastrophic death loss was not conclusively ascertained. The authors attributed the death of these cattle to type D botulism ("Lamziekte" or "lame sickness") on the basis of (1) findings in experimentally induced type D botulism (biological plausibility), (2) similarities of clinicopathologic findings with prior reports of naturally occurring type D botulism in cattle, some of which had occurred in that area (consistency), and (3) elimination of other possibilities (rule out). The rule out list appears in Table 12.4.

The high morbidity over a short period (11 days) suggested a point source consistent with contamination of feed or water. The suspected toxin could not be demonstrated by standard laboratory techniques or through feeding in the bone meal, concentrate fed, water, in gut contents, feces or serum of affected cattle. Rat carcasses, a potential source of clostridial toxin,

Table 12.5 Case fatality rates for *Salmonella* serotypes isolated from horses at the VMTH, UC Davis, 1971 to 1982

Salmonella Serotype	*No. of Cases*	*Case Fatality Rate (%)*
typhimurium var copenhagen	43	62.8
typhimurium	63	58.7
anatum	50	38.0
saint-paul	24	33.3
kottbus	14	21.4
Mixed*	12	41.7
Other	39	28.2

* At least one member of each mixed pair was of the listed serotypes.
Mean case fatality rate for all serotypes = 110 ÷ 245 x 100 = 44.9%.

Reprinted with permission from Carter, J.D., Hird, D.W., Farver, T.B., and Hjerpe, C.A. 1986. Salmonellosis in hospitalized horses: seasonality and case fatality rates. *J.A.V.M.A.* 188:163-167.

were not found in the feeding system. In light of the 7- to 17-day incubation period, the suspected feed could have been fed out by the time samples were drawn. Lack of illness in domestic ducks does not rule out water contaminated by decaying animal carcasses as a source of toxin. Birds are much less sensitive to type D *C. botulinum* toxin than cattle.

H. 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.

V. CASE STUDY

A. RISK FACTORS FOR SALMONELLOSIS IN HOSPITALIZED HORSES (HIRD ET AL, 1986)

In this case study multiple causes are ranked by strength of association.

1. Introduction

Earlier in the text a case study was presented describing the occurrence of 245 equine salmonellosis cases over an 11-year period in a VMTH (Carter et al, 1986; Chapter 11). A time series analysis revealed no significant overall increase or decrease in the rates (secular trend) over the 11-year period, but seasonal and cyclical variations were detected. There was no regular pattern in the cycles that would be useful for forecasting salmonellosis outbreaks at the VMTH.

Table 12.6 Case fatality rates for certain risk factors in horses developing nosocomial* salmonellosis at the VMTH, UC Davis, July 1971 to June 1982. The number of horses at risk during the interval was 14,330

Risk Factor	*No. of Times Diagnosed*	*No. of Cases*†	*Case Fatality Rate (%)*§
Impaction	1088	53	47.2
Exploratory laparotomy	449	49	55.1
Compromised blood supply to gastrointestinal tract	294	20	50.0
Minor locomotor problem	1513	16	31.3
Major respiratory tract disease	544	11	45.4
Miscellaneous infection outside of gastrointestinal tract	589	11	27.3
Enteritis	319	9	44.4
Major locomotor problem	548	8	62.5
Myopathy	84	6	66.7
Other**	1565	25	40.0

* Nosocomial salmonellosis cases are those horses from which the first specimen was obtained for bacteriologic analysis, or which developed clinical signs ≥72 hours after admission.

† Some horses were counted in several risk factor categories.

§Differences between rates were not significant.

** Fewer than five cases per risk factor.

Reprinted with permission from Carter, J.D., Hird, D.W., Farver, T.B., and Hjerpe, C.A. 1986. Salmonellosis in hospitalized horses: seasonality and case fatality rates. *J.A.V.M.A.* 188:163-167.

Eighteen serotypes were implicated. The overall case fatality rate was 44.9%. Although the case fatality rate for *S. typhimurium* and *S. typhimurium* var *copenhagen* infections was significantly greater than all other serotypes combined (Table 12.5), no significant differences were found among case fatality rates for different breeds, sexes, ages and other risk factors (Table 12.6).

2. Purpose of the Study

The objective of this follow-up case-control study was to identify risk factors for the development of nosocomial *Salmonella* infections in horses hospitalized at the VMTH.

3. Epidemiologic Methodology

Cases of nosocomial infection were defined as horses hospitalized at the VMTH between July 1, 1971 and June 30, 1982, from which *Salmonella* were obtained from specimens submitted 72 or more hours after admission. Horses from which *S. saint-paul* was isolated during the 1982 through March 1982 outbreak were excluded from the study. Preliminary identification of cases was accomplished by computer search of the Veterinary Medical Data Base records of the hospital. Additional cases were selected from the microbiology laboratory log book.

Two control groups were formed. The first consisted of the total hospitalized equine population for the 11-year period, excluding *Salmonella* cases. The second control group consisted

of randomly selected patients whose month of discharge coincided with the month in which a case developed. Equal numbers of case and control horses were used.

A variety of risk factors were recorded from the hospital records for each case and control, and coded for computer analysis. Descriptive statistics were calculated for non-categorical variables, and means for cases and controls were tested for significant differences, using Student's t-test.

a. *Bivariate Analysis*

Unadjusted odds ratios of *Salmonella* isolation following exposure to individual risk factors during hospitalization were calculated and tested for significance by the Chi-square test. Using this approach, odds of isolation from horses in one category were compared with odds for horses in all other categories combined.

b. *Multivariate Analysis*

Stepwise multiple logistic regression analysis was used to estimate the odds of *Salmonella* isolation on the basis of risk factors determined to be significant in the preceding analysis, while controlling for the possible confounding effects of other risk factors.

4. Assumptions Inherent in the Methodology

Since specimens for bacterial culture were not routinely taken from clinically normal horses at the hospital, the risk factors being evaluated should be considered risk factors for isolation of *Salmonella* rather than risk factors for infection. Although some controls were undoubtedly infected, their proportion was too small to have affected the conclusions of the study. Although the classification of an infection as nosocomial was arbitrary, the similarity of the authors' findings to those of previous studies was interpreted as supportive of the validity of their classification system. Initial selection of risk factors using unadjusted odds ratios reduced the number of confounders to adjust for when estimating adjusted odds ratios.

5. Basic Epidemiologic Findings

a. *Bivariate Analysis*

One hundred thirty-one cases of nosocomial salmonellosis were identified. Unadjusted odds of nosocomial salmonellosis for various intrinsic and extrinsic risk factors during hospitalization are shown in Table 12.7. Examination of unadjusted odds ratios revealed that 18 risk factors were significantly associated ($P < 0.01$) with isolation of *Salmonella* from nosocomial infections.

b. *Multivariate Analysis*

Logistic regression coefficients and their standard errors were used to calculate the adjusted odds ratios, with their corresponding 95% confidence intervals. Examination of odds ratios after adjustment for confounding variables reduced the number of risk factors from 18 to 3 (Table 12.8). Horses in which nasogastric tubes were passed (after controlling for amount of systemic antibiotic usage and admission because of colic) were at 2.85 greater risk of *Salmonella* isolation compared with horses in which tubes were not passed. Horses treated parenterally or both orally and parenterally were at 6.35 and 40.41 times greater risk of salmonellosis, respectively, than those not receiving such treatment. Horses admitted because of colic were 4.21 times as likely to have *Salmonella* isolated as those admitted for other reasons. Type of surgery, emergency admission, age, sex, breed and interactions were not believed to be important risk factors.

6. Conclusions and Measures Taken

Because bacteriologic culturing of nasogastric tubes and other materials had not been per-

Table 12.7 Unadjusted odds ratios* of *Salmonella* isolation for statistically significant ($P < 0.01$) risk factors among horses hospitalized during the period July 1971 through June 1982 at the VMTH, UC Davis

Risk Factor	*Percentage of Cases (n = 131)*	*Percentage of Controls (n = 131)*	*Unadjusted Odds of Isolation*
	Intrinsic (Host) Factors		
Breed			
Arabian	17.7	6.2	3.25
Thoroughbred	14.6	32.6	0.35
Presenting complaint			
Colic	46.6	9.9	7.91
Lameness	16.0	41.2	0.27
Emergency at admission	32.1	14.6	2.76
	Extrinsic Factors		
Procedures before isolation			
Nasogastric intubation	45.0	9.9	7.44
Intravenous catheter	90.0	51.1	8.67
Rectal palpation	59.5	20.6	5.67
Surgery performed	68.7	46.9	2.48
Anesthesia			
None	32.8	52.5	0.44
Inhalant	60.7	39.2	2.39
Type of surgery			
No surgery	32.6	53.1	0.43
Noninvasive	3.9	16.2	0.21
Invasive-abdominal	22.5	2.3	12.28
Enterotomy	17.1	2.3	8.70
Antibiotic administration			
None	11.9	48.8	0.14
Parenteral	74.6	49.6	2.99
Oral and parenteral	12.7	0.8	17.75

* Odds of isolation: horses in category versus horses in all other categories.

Reprinted with permission from Hird, D.W., Casebolt, D.B., Carter, J.D., Pappaioanou, M., and Hjerpe, C.A. 1986. Risk factors for salmonellosis in hospitalized horses. *J.A.V.M.A.* 188:173-177.

Table 12.8 Adjusted odds ratios of *Salmonella* isolation and 95% confidence intervals computed from logistic regression coefficients and their standard errors for risk factors significantly associated ($P < 0.05$) with isolation of *Salmonella* from horses hospitalized at the VMTH, UC Davis, from July 1971 through June 1982

Risk Factor	*Adjusted Odds Ratio**	*95% Confidence Interval†*
Nasogastric intubation		
Not performed§	1.00	
Performed	2.85	1.87 - 4.32
Antibiotic administration		
None§	1.00	
Oral	1.28	0.12 - 13.22
Parenteral	6.35	2.41 - 16.75
Oral and parenteral	40.41	7.09 - 230.32
Presenting complaint of colic		
Without complaint§	1.00	
With complaint	4.21	2.79 - 6.37

* Odds adjusted for all other variables in the equation.

† Confidence intervals not including 1 indicate statistical significance at the 5% level.

§ Reference category.

Reprinted with permission from Hird, D.W., Casebolt, D.B., Carter, J.D., Pappaioanou, M., and Hjerpe, C.A. 1986. Risk factors for salmonellosis in hospitalized horses. *J.A.V.M.A.* 188:173-177.

formed, it was not possible to determine whether the association between *Salmonella* isolation and nasogastric intubation meant that *Salmonella* was transmitted by the tube itself or by associated materials and procedures. Use of antibiotics may enhance growth of *Salmonella* through selective elimination of gut flora antagonistic to *Salmonella*. The identification of colic as a risk factor may mean that horses with intestinal disturbances were more susceptible to infection with *Salmonella*. However, clinicians suspecting this association may have been more likely to submit fecal specimens from patients with colic, thus biasing the results of this study.

VI. SUMMARY

Epidemiologic investigation of a disease outbreak of unknown etiology will usually incriminate a number of factors, or *determinants*, in the disease. Determinants of disease include both the etiologic agents 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.

The biological properties of agents, such as pathogenicity and virulence, strains and genetic drift, are primary determinants of the ability of an agent to cause disease. 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. 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.

Populations also differ in susceptibility. Resistance in populations is called *herd 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 previous natural exposure or immunization. *Acquired herd immunity* results from the development of protective immunity in a population after natural exposure or immunization. Increased herd immunity has the effect of limiting the spread of diseases by reducing the proportion of *effective contacts*, e.g., contacts between infected and susceptible animals that result in transmission of a disease agent. As a result, the reproductive rate for a disease agent may fall below that required for its maintenance in the population, leading to its eventual eradication. It follows that the higher the *intrinsic reproductive rate* (R_0) of a disease organism, the higher the level of herd immunity that must be achieved for its eradication.

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 quantitate and manipulate.

A number of criteria may be applied to test the strength of a presumed cause-effect relationship when Koch's postulates cannot be used. These include (1) the strength of the study design, (2) demonstration of a temporal relationship between a hypothesized cause and effect, (3) the strength of the association between a presumed causal factor and outcome, (4) demonstration of a dose-response relationship between a suspected cause and effect, (5) biological plausibility of the presumed cause-effect relationship, (6) consistency of findings in studies conducted in different settings and with different patients, (7) elimination of other possibilities on the rule-out list and (8) demonstration of a reversible association between presumed cause and effect.

Chapter 13

SOURCE AND TRANSMISSION OF DISEASE AGENTS

I. SOURCES OF INFECTION

A. IATROGENIC INFECTIONS

Some of the risk factors for nosocomial *Salmonella* infections in horses discussed in the preceding chapter implicated medical procedures in the transmission of bacteria to hospitalized patients. Iatrogenic illnesses, e.g., 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, or the inappropriate use of particular therapeutic regimens, are examples of iatrogenic illnesses.

In some cases, as when attenuated agents are injected for vaccinal purposes, 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 the vaccine as the source of the disease agent. The recent availability of tools for the molecular characterization of microorganisms has given birth to a new branch of epidemiology – *molecular epidemiology.*

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: Extensive use of modified-live (ML) infectious bovine rhinotracheitis virus (IBRV) vaccine can reduce the frequency of infectious bovine rhinotracheitis (IBR). Although ML IBRV vaccines are widely used, their safety has been questioned. The laboratory techniques used to date, such as serologic testing, are unable to differentiate vaccinal virus from field viral isolates. A study was conducted by personnel of the National Veterinary Services Laboratories, Ames, IA to (1) characterize the restriction endonuclease analysis (REA) patterns of the 14 ML IBRV vaccine strains currently licensed, and (2) evaluate the role of vaccinal virus in field epidemics of IBR. Viral DNA from isolates obtained from six field samples of IBRV (Colorado, 1; West Virginia, 1; Wisconsin, 3; South Dakota, 1) were digested with restriction endonucleases, and patterns were compared with vaccinal virus. Animals from which field samples were obtained had been vaccinated with ML IBRV vaccine before an epizootic of IBR occurred in the herds. In two of the six field samples, DNA REA patterns from the isolates were indistinguishable from the pattern for the vaccinal viruses used. In the remaining four field samples, DNA REA patterns of the IBRV from isolates were different from those of the vaccinal virus. The authors caution that the most conclusive results are obtained only when the REA patterns are distinctly different, not when they are the same (Whetstone et al, 1986). In other words, it is easier to prove that virus strains are different than prove that they are the same.

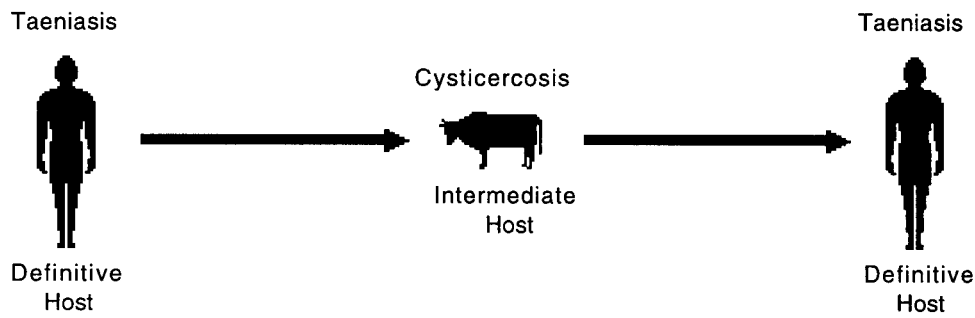


Figure 13.1 Taeniasis and cycticercosis (*Taenia saginata*) – transmission cycle. (Reprinted with permission from Acha, P.N. and Szyfres, B. 1980. *Zoonoses and Communicable Diseases Common to Man and Animals.* Pan American Health Organization. Washington, D.C. 700 pp. Copyright Pan American Health Organization, Washington, D.C.)

B. ANIMAL RESERVOIRS

Animal reservoirs of disease agents include (1) carrier animals, animals (and human beings) with inapparent infections that are also transmitters (or potential transmitters) of the infectious agent, and (2) intermediate hosts and vectors. Amplifying hosts may play a role in creating conditions favorable for epidemics of a disease.

Animals that have been exposed to an agent may become carriers. *Incubatory carriers* are capable of serving 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 the signs and symptoms of disease have disappeared, i.e., recovery. This is seen with many parasitic infections caused by protozoa and helminths.

We tend to look at nature anthropocentrically, i.e., regarding human beings as the central fact or final aim of the universe. In the case of zoonotic diseases, this means viewing animals as a source of infection for human beings. In some cases, human beings may be an important source of infection for other animals.

EXAMPLE: Cysticercosis is a disease of cattle caused by encysted larvae of the cestode *Taenia saginata.* Cattle are intermediate hosts while humans are the definitive host and source of infection for cattle (Figure 13.1).

From January to March 1981, 37 slaughter cattle from a single Ohio feeding operation were determined, at postmortem inspection, to be infected with *T. saginata* cysticerci. A subsequent outbreak on the same farm in March 1983 involved seven slaughter cattle. By multiplying the prevalence rate of cysticercosis detected at federally inspected plants in Ohio by the number of cattle slaughtered at the Ohio Department of Agriculture (ODA) inspected plants, eight cases per year would have been expected in ODA-inspected plants. Applying this same prevalence rate to the total number of cattle slaughtered from this farm in 1980, the expected number of cases was 0.005. The observed number, 37 cases, was 7400 times greater than expected and therefore constituted an outbreak.

An epidemiologic investigation was conducted of possible sources of the *T. saginata* ova; these included (1) leakage of raw sewage onto the pasture after a flood in 1980, (2) municipal sewage sludge application on the farm, (3) defecation in feed or water by farm workers and (4) infection of cattle before arriving at the farm.

The farm consisted of approximately 243 hectares (1 hectare = approximately 2.5 acres) with 162 hectares for cropland and 81 hectares for pasture. Corn and hay were the only crops raised on this farm. A municipal sewage treatment plant was adjacent to the northeast corner

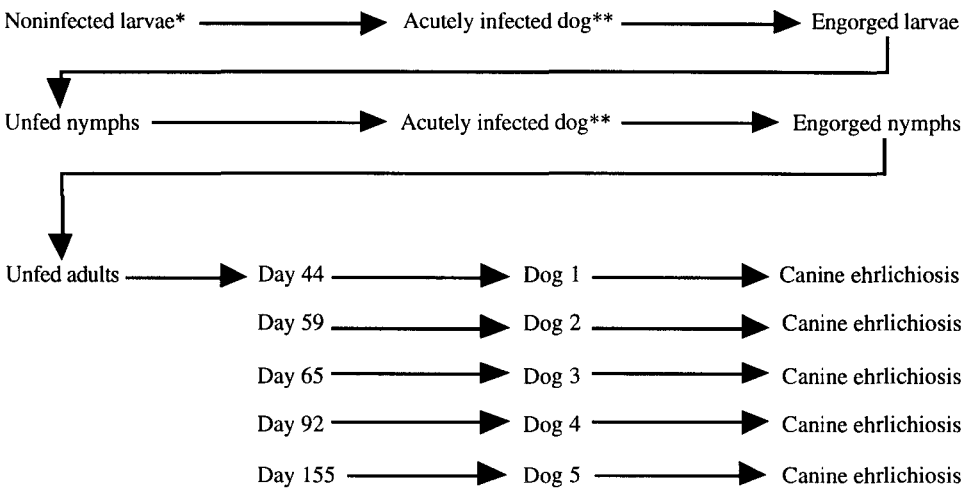


Figure 13.2 Evaluation of unfed adult *Rhipicephalus sanguineus* ticks, which fed as larvae and nymphs on acutely infected dogs, as reservoirs of *Ehrlichia canis*. (Reprinted with permission from Lewis, G.E., Jr., Ristic, M., Smith, R.D., Lincoln, T., and Stephenson, E.H. 1977. The brown dog tick *Rhipicephalus sanguineus* and the dog as experimental hosts of *Ehrlichia canis*. *Am. J. Vet. Res.* 38:1953-1955.)
* Unfed *R. sanguineus* larvae, maintained for two previous generations on normal dogs.
** Acute canine ehrlichiosis: rectal temperature $\geq 39.2^{o}C$, parasites in peripheral blood monocytes and severe thrombocytopenia.

of the farm, downstream from a 5- to 10-m wide creek that ran through pastures grazed by affected cattle. Following the creek was a sewer line that terminated at the sewage plant. There were nine manholes, covered with loose fitting tops, along the sewer line in the pasture. These manholes were elevated approximately 30 cm above the pasture. The cattle had access to these manholes. On June 28, 1980 heavy rainfall occurred and much of the pasture and some croplands were flooded for approximately 4 to 5 days.

The farm had received applications of municipal sewage sludge intermittently for the last 20 years. The sludge originated only from the adjacent sewage treatment plant. During 1980 and 1982 (preceding the 1981 and 1983 cysticercosis outbreaks), sludge was applied to pastures.

Temporal and spatial observations implicated raw sewage contamination of pastures (from flooding) as the most likely source of infection in the 1981 outbreak. The outbreak in 1983 was more likely associated with sludge application. The possibility of an infected worker exposing the cattle to infected feces was not excluded as a possible source (Fertig and Dorn, 1985).

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.

EXAMPLE: *Ehrlichia canis*, the etiologic agent of canine ehrlichiosis, is a tick-borne rickettsia that can persist in the blood of infected dogs for periods of time that far exceed the lifespan of the tick vector, *Rhipicephalus sanguineus*, the brown dog tick. Notwithstanding, experimental studies revealed that the period of infectivity of the dog for the tick is restricted to the febrile phase of infection, which does not exceed 2 weeks. The tick appears to remain infective for life (Figure 13.2).

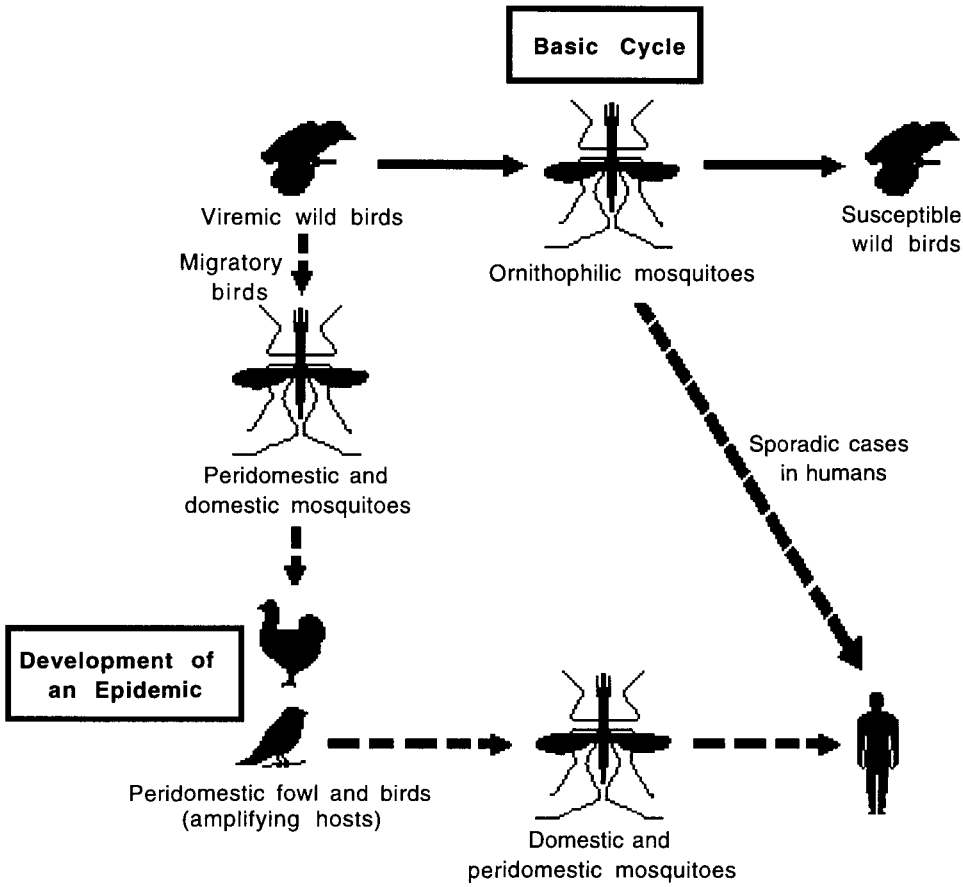


Figure 13.3 St. Louis encephalitis – probable cycle of virus and role of amplifying hosts in epidemics. (Reprinted with permission from Acha, P.N. and Szyfres, B. 1980. *Zoonoses and Communicable Diseases Common to Man and Animals.* Pan American Health Organization. Washington, D.C. 700 pp. Copyright Pan American Health Organization, Washington, D.C.)

Adult *R. sanguineus* ticks efficiently transmitted *E. canis* to susceptible dogs for 155 days after detachment as engorged nymphs from a dog in the acute phase of ehrlichiosis. Adult ticks that had similarly engorged on a dog in the chronic phase of ehrlichiosis failed to transmit *E. canis* to susceptible dogs. Infected but unfed adult ticks may thus be of greater importance than the chronically infected carrier dog as a natural reservoir of *E. canis* (Lewis et al, 1977).

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 and provides a source for epidemics in humans or domestic animals. St. Louis encephalitis provides an excellent example (Figure 13.3).

The basic cycle of the infection involves wild birds and ornithophilic mosquitoes. Peridomestic birds and domestic fowl serve as amplifiers of the virus. That, together with increased density of the human population, creates the conditions necessary for epidemics. How the virus gets into urban areas is not yet established, though it is suspected that migratory wild birds are responsible.

C. 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, nontransmissible disease agent. Infection results from inhalation of airborne conidia, which are produced during growth of organisms in the soil. See the next section for a further discussion of transmissible versus nontransmissible 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. From the standpoint of control it may be unwise to restrict our view to only the immediate source of infection. Consider the following example.

EXAMPLE: In the 3-month period from October 17, 1985 to January 9, 1986, 44 episodes of pyoderma occurred among 32 workers in an Oregon meat-packing plant. Most of the 44 reports involved impetigo-like lesions (pustules) on the hand, wrist and forearm, but six episodes of cellulitis (inflammation of the cellular and subcutaneous tissue) and two of lymphangitis (inflammation of lymphatic vessels) were also reported. The same epidemic strain of Group-A, -B hemolytic *Streptococcus* (GAS) isolated from skin lesions was also isolated from meat in the plant. The attack rate for boners and killers was 74%, compared with 13% for workers who were never involved in killing or boning (relative risk = 5.7; 95% confidence interval = 2.9 - 11.3). The epidemic investigation suggested that though the infection was acquired from the environment, meat was a vehicle of transmission of GAS between workers, probably after initial contamination by an infected human. Knife use was probably the significant risk shared by killers and boners versus other meat workers.

Recommendations to the meat-packing plant included an increased emphasis on worker safety; an increased emphasis on worker hygiene, e.g., covering skin lacerations; removal of workers with untreated skin infections from the meat-processing line; and improved surveillance of skin injuries and infections, including modifying sick-leave benefits to encourage reporting (CDC, 1986b).

II. TRANSMISSION

A. 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 route 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 (Anderson and

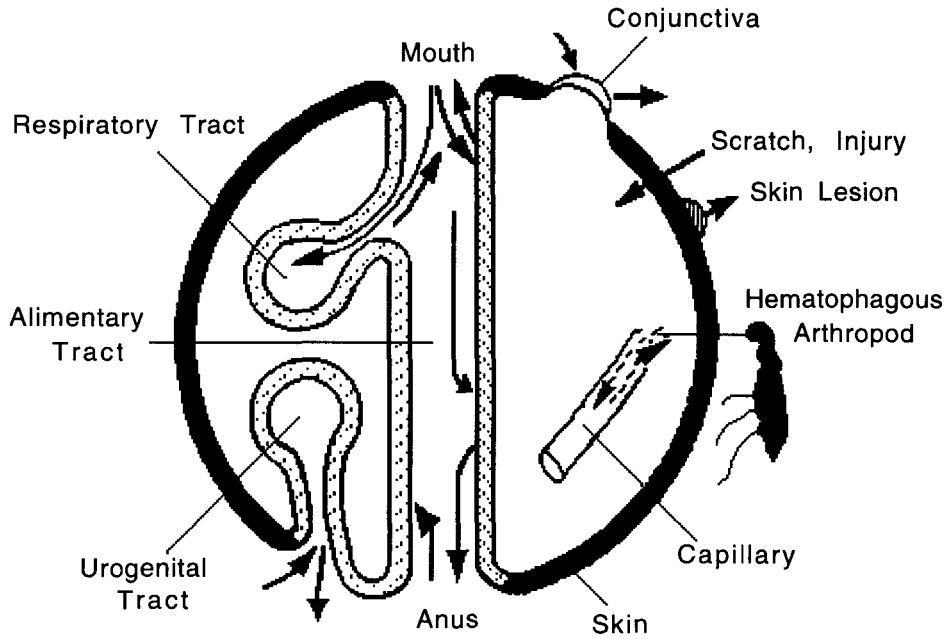


Figure 13.4 Diagram illustrating the routes of exit and entry of infectious agents in vertebrate animals. (Source: Anderson, R.M. and May, R.M. 1982. *Population Biology of Infectious Diseases.* Springer-Verlag, New York. 315 pp.)

May, 1982) (Figure 13.4). Mode of spread or dissemination refers to how a disease agent is spread from one geographic area to another.

B. TRANSMISSIBLE VERSUS NONTRANSMISSIBLE DISEASES

Diseases are broadly classified as *transmissible* (communicable) or *nontransmissible.* Transmissible disease may be due to a specific infectious agent or its toxic products (such as Type D botulism toxin, see Table 12.4), 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.

Nontransmissible diseases may be caused by infectious or noninfectious 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. Noninfectious agents include poisons and environmental toxins, immunologic and metabolic mechanisms, nutritional deficiencies and functional defects (such as congenital anomalies).

Introduction into the herd of an animal afflicted with a nontransmissible 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.

Contact with diseased animals is always viewed with some degree of apprehension. Practically speaking, introduction into the herd of an animal afflicted with a nontransmissible

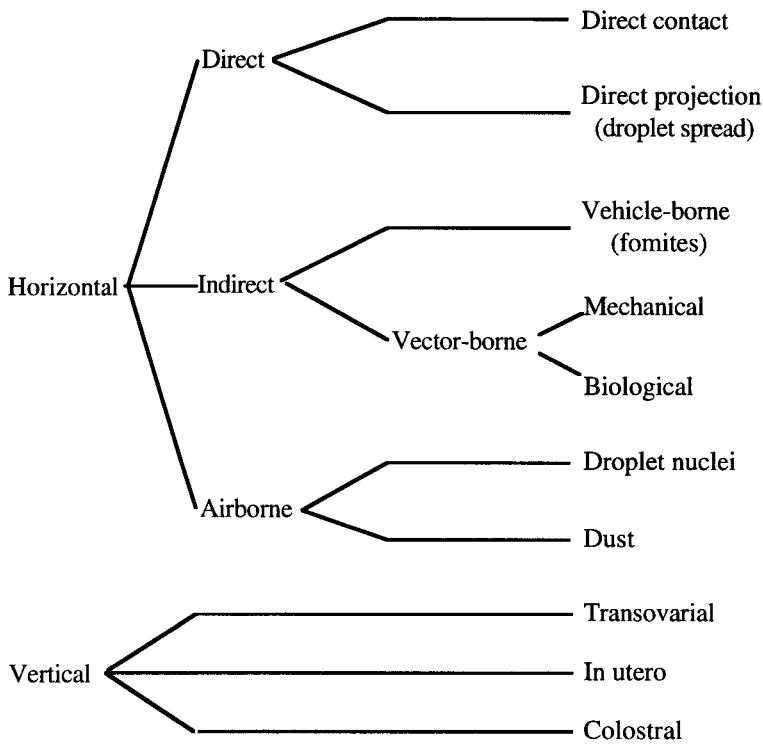


Figure 13.5 Transmission of infectious agents.

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.

III. MODES OF TRANSMISSION

Transmission may occur *horizontally* by transmission of an infectious agent 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 13.5 and described in the following sections (Schwabe et al, 1977; Benenson, 1985).

A. HORIZONTAL TRANSMISSION

Horizontal transmission describes the transmission of a disease agent among contemporaries. Modes of horizontal transmission may be direct, indirect or airborne.

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. The so-called *fecal-oral* mode of transmission is also direct since it requires direct contact between the susceptible and infected individual, whose skin or hairs are contaminated with feces. A second mode of direct transmission is through *direct projection*, where atomized droplets are sprayed onto the conjunctiva or mucous membranes of the eye,

Table 13.1 Febrile and serologic responses, and occurrence of abortion, in mares exposed to mares bred to stallions infected with equine viral arteritis virus

Horse No.	*Date Exposed*	*Initial Febrile Response (Days Post-exposure)*	*Seroconversion (Days Post-exposure)*	*Occurrence of Abortion*		
				Postexposure Day	*Day After Febrile Response*	*Fetal Age (Days)*
1	12/28/84	27	28	41	14	281
2	"	25	28	35	10	224
3	"	25	28	34	9	288
4	"	22	25	37	15	312
5	"	28	28	57	29	288
6	"	NR	28	NR	NR	NR
7	"	18	24	24	6	251
8	"	20	27	29	9	216
9	"	3	17	23	20	258
10	"	18	28	25	7	337
11	"	17	26	40	23	312
12	"	15	27	NR	NR	NR
13	"	8	20	NR	NR	NR
14	"	NR	UD	NR	NR	NR
T*	12/27/84	4	11			
U*	"	2	8			
V*	"	2	8			
W*	"	1	8			

*Mares bred to infected stallions on 12/27/84. On 12/28/84 these mares were used for contact exposure to mares 1 through 14.
NR = no response; UD = undetermined.

Reprinted with permission from Cole, J.R., Hall, R.F., Gosser, H.S., Hendricks, J.B., Pursell, A.R., Senne, D.A., Pearson, J.E., and Gipson, C.A. 1986. Transmissibility and abortogenic effect of equine viral arteritis in mares. *J.A.V.M.A.* 189:769-771.

nose or mouth during coughing or sneezing. Direct projection, also known as *droplet spread*, is usually limited to a distance of 1 m or less.

EXAMPLE: Equine viral arteritis was first identified in the United States in 1953 after an outbreak of abortions in mares and generalized illness in horses. The disease is characterized by fever, leukopenia, stiffness of gait, edema of the limbs, swelling around the eyes with conjunctivitis and lacrimation and a serous nasal discharge. The causative agent is an RNA virus in the family Togaviridae. A 1984 outbreak among Kentucky Thoroughbreds prompted a request by the Thoroughbred industry and the USDA to determine (1) whether pregnant mares (4-7 months gestation) could be infected with equine viral arteritis virus via contact with infected mares, and (2) whether this infection would cause abortion (Cole et al, 1986).

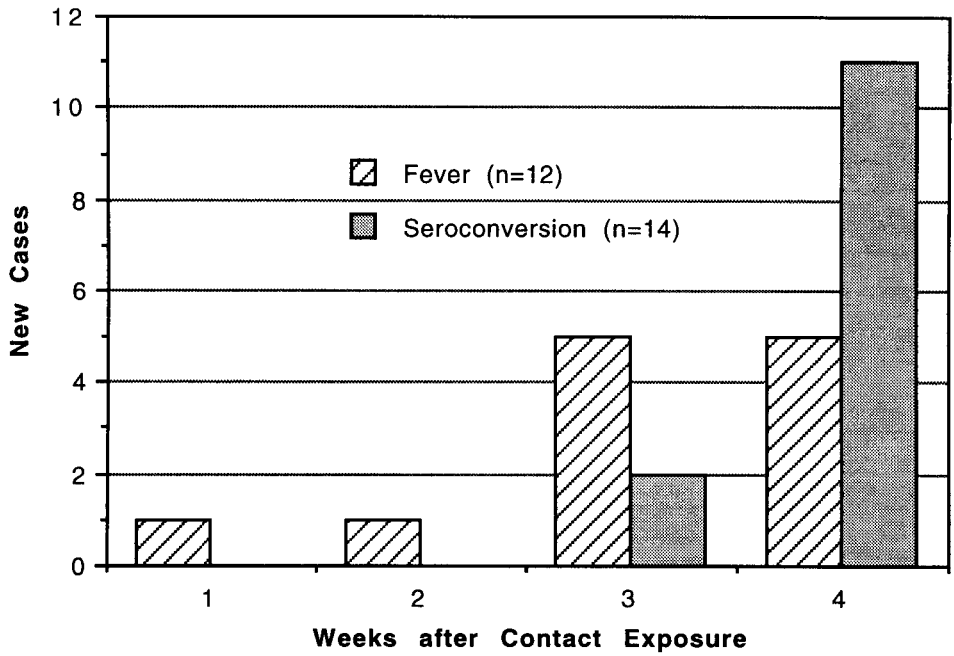


Figure 13.6 Transmission of equine viral arteritis – onset of fever and seroconversion among contact mares after commingling with mares exposed by breeding to infected stallions. (Source of data: Cole, J.R., Hall, R.F., Gosser, H.S., Hendricks, J.B., Pursell, A.R., Senne, D.A., Pearson, J.E., and Gipson, C.A. 1986. Transmissibility and abortogenic effect of equine viral arteritis in mares. *J.A.V.M.A.* 189:769-771.)

Fourteen equine viral arteritis-negative pregnant mares (recipients) were divided into two groups of seven each and each group commingled with two mares (donors) that had been bred the previous day to naturally infected stallions. The mares were confined to an open corral that was divided into two pastures by wire fencing. Rectal temperature was recorded daily and blood samples collected a minimum of every 7 days.

Both donors and 12 of 14 recipient mares developed fever. All mares seroconverted within 1 month of exposure. Clinical signs of the disease in donors included anorexia, conjunctivitis and nasal discharge (4), and lameness (2). Recipient mares developed anorexia (8), lameness (4), conjunctivitis (2) and nasal discharge (1).

Table 13.1 shows the relationship of the febrile response, seroconversion and time of abortion to the time of exposure. Onset of fever and seroconversion from the time of commingling are depicted in Figure 13.6. The results demonstrate how efficiently some disease agents can be transmitted through direct contact. The protracted onset of signs of disease suggest either a propagating epidemic or highly variable incubation period.

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.

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.

a. *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* (Halpin, 1975). 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.

EXAMPLE: Bovine leukosis virus (BLV) is an exogenous retrovirus of cattle that persists for the life of infected animals in bone marrow-derived lymphocytes. As no virus is present in nasal secretion, saliva, urine or semen, except when those fluids are contaminated by blood or cellular exudate, it is postulated that most cattle become infected by exposure to virus-infected lymphocytes, rather than cell-free virus. It follows that procedures that result in the transmission of whole blood containing BLV-infected lymphocytes from animal to animal are important in the spread of BLV in cattle. This hypothesis is supported by reports attributing clusters of leukosis in cattle to piroplasmosis (babesiosis) vaccination with blood from leukotic cattle. Blood-contaminated dehorning instruments have also been implicated in the spread of BLV in dairy calves. A study was conducted to determine whether small volumes of whole blood, simulating farm practices, resulted in viral transmission.

Four calves were given 10 µl each of whole blood from a BLV-carrier cow by intramuscular, intravenous, subcutaneous or intradermal routes. An additional four calves were given 1 µl of blood from the same donor cow by the previously mentioned routes. The first four calves had all seroconverted (to BLV-positive) within 8 weeks of exposure, and the latter four calves within 14 weeks. Although the infectious dose (actual number of viral units transmitted) was not known, the results support the hypothesis that the use of common needles in vaccinations or parenteral injections enhances the spread of BLV to susceptible animals in the population (Evermann et al, 1986). In this case contaminated needles, dehorning instruments, ear punches, etc. would be considered vehicles or fomites.

EXAMPLE: A 37-year-old man became ill with signs and symptoms compatible with leptospirosis. Three days later, he entered a hospital with a temperature of 103.6°F, slightly abnormal liver function tests, leukopenia and mild anemia. He was started on tetracycline, and 12 hours later his symptoms cleared. Thirty days later he again had fever, headache and myalgia. This time his symptoms were accompanied by bilateral orchitis. He was given oral ampicillin, and 3 days later his symptoms cleared. Paired serum samples collected after initial onset of disease showed increasing titers for *Leptospira ballum* in the microscopic agglutination test.

The patient had purchased two white mice at a local pet shop approximately 3 months before the initial illness. Both mice were sacrificed and found to have nephritis. Spirochetes were isolated from their kidneys. When inoculated into guinea pigs, the spirochetes caused a diagnostic titer rise in the guinea pigs for *L. ballum*. A mouse obtained from the mouse

colony that was the source of these animals for the pet shop was also found to harbor *L. ballum* in its kidneys. Sera obtained from the patient's wife and three daughters, as well as the man and woman who owned the mouse colony, were all negative for leptospiral antibodies.

Because the patient had virtually no contact with the pet mice, the route of infection was uncertain. The patient speculated that one of his daughters, after an argument, had used his toothbrush to clean the mouse cage (Friedmann et al, 1971).

b. *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. 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.

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 to be airborne. Airborne transmission may be effected 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.

B. 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, rickettsiae and protozoa.

IV. 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 *intrinsic* (or *basic*) *reproductive rate* (R_0), a dimensionless parameter defined as the average number of secondary cases of infection to which one primary case gives rise throughout its infectious period if introduced into a defined population consisting solely of susceptible individuals (Anderson and May, 1982).

One way of expressing communicability is the intrinsic (or basic) reproductive rate (R_0), a dimensionless parameter defined as the average number of secondary cases of infection to which one primary case gives rise throughout its infectious period if introduced into a defined population consisting solely of susceptible individuals.

Based on the few epidemiologic studies that have attempted to measure R_0, vector-borne infections (e.g., malaria and filariasis) appear to attain higher maximum values when compared with other directly and indirectly transmitted infections (e.g., measles, hookworm, ascariasis) (Anderson and May, 1982; see Table 12.1). The value of R_0 is 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.

A. AGENT FACTORS

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* 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 (Benenson, 1985). 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.

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.

B. HOST FACTORS

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. The term lousy refers to the propensity of certain individuals to develop heavy louse infestations, particularly in the winter. In cattle operations it is recommended that these animals be eliminated from the herd, rather than treated.

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.

C. ENVIRONMENTAL FACTORS

1. Particle Diameter

a. *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 time of suspension is less than three seconds. Their flight range is restricted to about 1 m or less. Droplet spread can be effectively reduced through use of a face mask and by reducing crowding among animals (Schwabe et al, 1977).

b. *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.

c. *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.

2. Microclimate

Among environmental factors, desiccation plays a major role in reducing transmissibility of infectious agents. Levine (1963, 1965) used bioclimatographs to predict the effect of climate on the epidemiology of sheep nematodes. Climatographs are graphs in which total precipitation is plotted against mean temperature for each month, and the resultant points are joined in a closed curve. Bioclimatographs are climatographs on which lines indicating the limits of climatic conditions most favorable for propagation of life, in this case free-living stages of ruminant nematodes, have been superimposed.

A climatograph for Urbana, IL based on meteorologic data from 1903 to 1954 is presented in Figure 13.7. Optimal conditions for pasture transmission of *Haemonchus*, *Trichostrongylus* and *Ostertagia* are superimposed. The resulting graph is a bioclimatograph. Urbana is suitable for *Haemonchus* pasture transmission throughout the summer and for *Trichostrongylus* and *Ostertagia* pasture transmission only during the spring and fall. *Haemonchus* is therefore a more important genus in the region, although *Trichostrongylus* and *Ostertagia* may occur (Levine, 1965).

The suitability of other regions for these three parasites can be compared by substituting monthly temperature and precipitation data for the Urbana data. The optimum condition lines for each parasite remain unchanged.

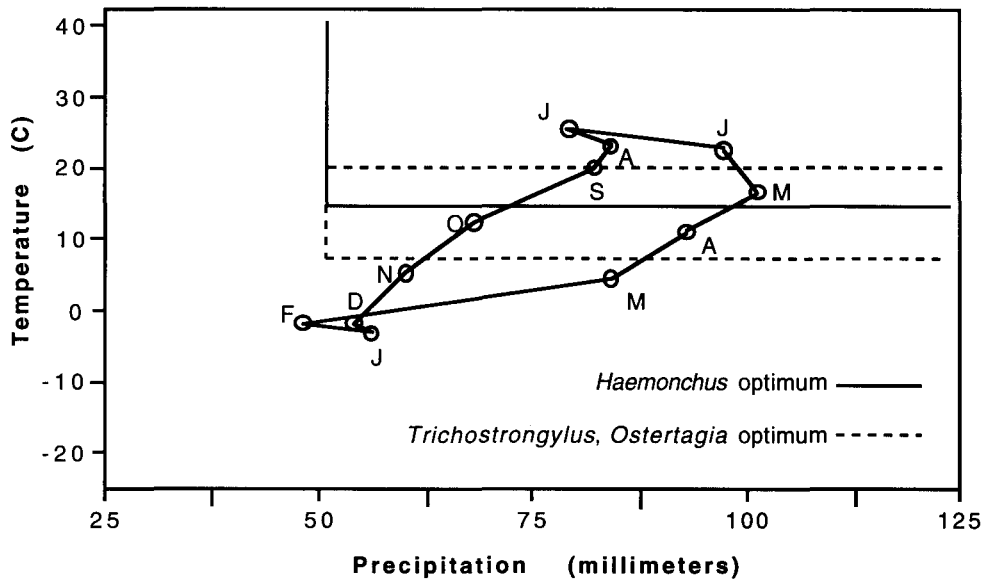


Figure 13.7 A bioclimatograph depicting months during which optimal conditions for pasture transmission of *Haemonchus*, *Trichostrongylus*, and *Ostertagia* occur in Urbana, Illinois based on climatic data from 1903 to 1954. Letters on the graphs are first letters of names of the months. (Adapted with permission from Levine, N.D. 1965. Bioclimatographs, evapotranspiration, soil moisture data and the free-living stages of ruminant nematodes and other disease agents. *Theoretical Questions of Natural Foci of Diseases,* B. Rosicky and K. Heyberger (eds), pp. 455-461. Czechoslovak Academy of Sciences.)

V. CASE STUDIES

A. TRICHINOSIS IN A HERD OF SWINE - CANNIBALISM AS A MAJOR MODE OF TRANSMISSION (HANBURY ET AL, 1986)

B. EPIDEMIOLOGIC FINDINGS ON A SWINE FARM WITH ENZOOTIC TOXOPLASMOSIS (DUBEY ET AL, 1986)

The source and mode of transmission of an endemic disease are determined.

1. Introduction

Trichinosis is a disease of wild and domestic animals accidentally transmitted to humans by the ingestion of meat or meat products. The etiologic agent, *Trichinella spiralis*, is a small filiform nematode, which in the adult stage lives a few weeks in the small intestine of a large number of mammalian species. In the larval state it forms a cyst in the musculature of these hosts, where it can remain viable for long periods (Acha and Szyfres, 1980).

Epidemiologically, two cycles can be distinguished: the domestic (synanthropic) and wildlife (sylvatic) cycles. The domestic and peridomestic cycles (Figure 13.8) center around the pig and include other animals such as dogs, cats and rats. The parasite is transmitted from pig to pig, mainly by the ingestion of garbage that contains muscle fibers of swine origin.

Dogs, cats and rats are infected from the same sources as pigs and are included in the cycle, but their epidemiologic role is secondary. Humans are an accidental host in whom the parasite does not find an appropriate exit to continue the cycle. The wildlife cycle is independent of the domestic cycle. Wild carnivores are the main reservoirs and the primary hosts of *T. spi-*

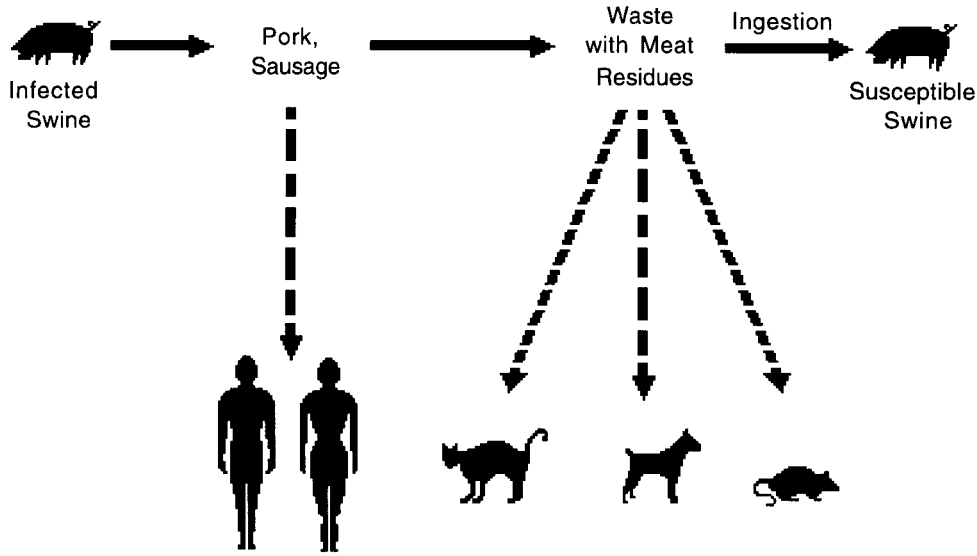


Figure 13.8 Trichinosis – synanthropic transmission cycle. (Reprinted with permission from Acha, P.N. and Szyfres, B. 1980. *Zoonoses and Communicable Diseases Common to Man and Animals.* Pan American Health Organization. Washington, D.C. 700 pp. Copyright Pan American Health Organization, Washington, D.C.)

ralis. The chief mode of transmission is by the consumption of carrion, generally consisting of older animals, which are the most intensely parasitized.

Trichinosis is a serious economic burden for the American swine industry (Hanbury et al, 1986). Control strategies include the development of serologic and meat inspection methods suitable for slaughterhouses, and research on carcass irradiation to inactivate larvae of *T. spiralis* in muscle. Critical information on important sources of *T. spiralis* for domestic swine in the United States is surprisingly scarce. Improperly cooked garbage has been considered a major source, but little attention has been given to other modes of transmission.

The advent of laws requiring the cooking of garbage appears to have decreased transmission of *T. spiralis* via the feeding of uncooked garbage. The importance of other possible sources such as commensal rats has been somewhat controversial. The importance of sylvatic trichinosis in the epidemiology of trichinosis in domestic swine is uncertain; evidence of sylvatic trichinosis is indirect or anecdotal, although experimental data indicate that the potential for transfer of *T. spiralis* from some fur-bearing species is high. Cannibalism of swine carcasses may also be a source of trichinosis.

Serologic surveys have indicated that about one third of pigs worldwide have been exposed to *Toxoplasma gondii* infection. In swine, several epidemiologic aspects of toxoplasmosis overlap with those of trichinosis, as can be predicted from similarities in their life cycles (Figure 13.9). Transplacental infection can develop in pigs, but probably is not a common route of *T. gondii* transmission. Soil, earthworms, feed or water contaminated with oocysts or tissue cysts appear to be the main sources of infection for pigs.

2. Purpose of the Study

From April 1973 to November 1983, 114 (60%) of 189 pigs from a farm in eastern Illinois were found to be infected with *T. spiralis*, as determined by digestion of diaphragmatic muscle (Table 13.2). The prevalence of *T. spiralis* infection in rats trapped on the farm was 45.5% in 1973, 33.3% 1975 and 7.5% in 1976.

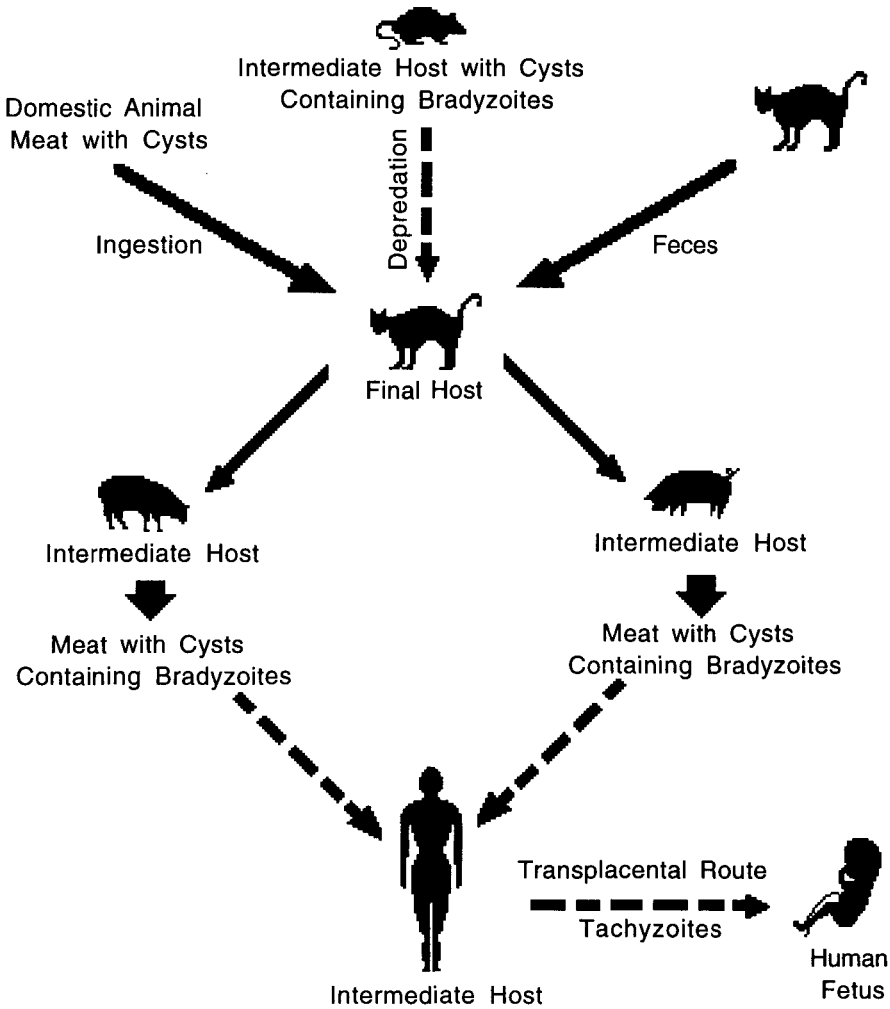


Figure 13.9 Toxoplasmosis – transmission to domestic animals and humans. (Reprinted with permission from Acha, P.N. and Szyfres, B. 1980. *Zoonoses and Communicable Diseases Common to Man and Animals.* Pan American Health Organization. Washington, DC. 700 pp. Copyright Pan American Health Organization, Washington, D.C.)

The purpose of the present study was to determine the relative importance of cannibalism, rodents or other wild animals in the epidemiology of *T. spiralis* infection in the herd. Prospective seroepidemiologic studies were also conducted using a modified agglutination test for detection of *Toxoplasma* antibody.

3. Epidemiologic Methodology

a. *Infected Swine Herd*

The present investigation was performed from January 1984 to March 1985. The farm was approximately 81 hectares of rolling pasture and woodlot. The swine herd was confined to six lots, some encompassing woods and a stream. The owner had 15 to 20 horses, 80 to 115 cattle, two goats, some geese, and about 1000 hogs; most of the hogs were sows and boars. The

Table 13.2 Results of muscle specimen digestion by use of ELISA for *Trichinella spiralis* in pigs on a farm in eastern Illinois from April 1973 to March 1985

Date (mo/yr)	*Pigs*	*No. of Specimens Evaluated**	*No. of Specimens Infected*	*Percent Infected†*
		Prior to study period		
4/73	Market	2	2	100
6/73	Market	3	3	100
7/73	Market	128	74	57
7/74	Sow	53	33	62
10/74	Market	2	2	100
11/83	Market	1	1	100
		Study period		
1/84	Dead	2	0	0
4/84	Dead	3	1	33
5/84	Market	2	1	50
8/84	Market	1	1	100
8/84	Dead	2	0	0
9/84	Sow	1	1	100
11/84	Dead	1	0	0
12/84	Dead	25	4	16
12/84	Young	138	28	20
1/85	Dead	4	0	0
3/85	Dead	4	2	50
Total (muscle digestion only)		234	124	
Total (ELISA only)		138	28	

*Pigs were evaluated by muscle specimen digestion, except for young pigs evaluated on 12/84, which were evaluated by ELISA.

†(No. of pigs infected with *T. spiralis*/No. of pigs evaluated) x 100.

Mean percentage of specimens infected that were evaluated by use of muscle digestion (n = 234) = 52.99%.

Mean percentage of specimens infected that were evaluated by use of ELISA (138) = 20.29%.

Sow = marketed at 180 kg.

Dead = carcasses found on farm (various weights).

Young = feeder and growing pigs up to 66 kg.

Reprinted with permission from Hanbury, R.D., Doby, P.B., Miller, H.O., and Murrell, K.D. 1986. Trichinosis in a herd of swine: cannibalism as a major mode of transmission. *J.A.V.M.A.* 188:1155-1159.

primary feed for the swine was waste grain from a food processing establishment. Raw garbage was not fed.

Most of the swine were born and raised on the farm. Young pigs, sows and boars (some up to 200 kg) frequently were kept together in the same lot, with as many as 400 pigs in a 24-hectare lot. Sows farrowed in various open lots or in any area in which they could gain access, including feed storage areas and barns.

The owner of the farm agreed to cooperate in the study conducted by personnel of the Division of Meat, Poultry, and Livestock Inspection (MPLI), Illinois Department of Agriculture. Because poor management (delayed removal of dead animals) was a potential factor in the transmission of *T. spiralis*, specific instructions on management improvement were given to the owner. To ensure compliance with these regulations and to maintain the integrity of the experimental design, daily visits to the farm were made.

b. *Serologic and Parasitologic Evaluations*

Blood was drawn from each hog and tested for antibody to *T. spiralis* with an ELISA test. Muscle specimens from tongue and diaphragm were collected from hogs at a local abattoir or from those found dead on the farm. Muscle specimens from tracer pigs were collected at the University of Illinois' Meat Sciences Department abattoir. All were examined by a digestion technique for *T. spiralis* larvae and level of infection expressed as larvae per gram (LPG).

c. *Longitudinal Experiment with Tracer Pigs*

Tagged tracer pigs, weighing 20 to 25 kg and serologically negative for antibody to *T. spiralis*, were used to evaluate sources and modes of transmission of *T. spiralis* on the farm. An area of the farm that had previously been occupied by infected pigs was fenced off to prevent entry of the owner's pigs beginning 3 weeks before the experiment. Thirty-nine tracer pigs (group 1) were housed (lot 1) and monitored so as to prevent cannibalism and exposure to rats or other wild animals. Lot 1 consisted of a double-fenced unit erected on a concrete platform. The outer fence was set approximately 30 cm into the ground and brought out approximately 30 cm to prevent access by burrowing animals. It was also electrified to prevent animals from climbing the fence.

Forty-five other pigs (group 2) were penned separately (lot 2) and monitored only to minimize cannibalism. Lot 2 consisted of a fenced section completely surrounding lot 1 and was adjacent to lots containing the owner's hogs. Lots 1 and 2 were monitored daily by MPLI personnel. Management procedures were performed by regular farm workers in a manner consistent with those used for swine in surrounding pens.

Tracer pigs were monitored daily during the 5 to 6 months required to reach market weight. Blood was collected from tracer pigs every 2 months, and those that became serologically positive were removed. Infection of the pigs was confirmed by digestion of muscle specimens after slaughter. At week 20 of the study, 12 of the 45 group 2 pigs were removed from lot 2 and placed in an isolation pen for a controlled hog cannibalism experiment.

d. *Hog Cannibalism Experiment*

The initial plan called for commingling tracer pigs with the owner's herd to serve as a group having potential exposure to cannibalism, rodents and other wild animals. The aggressive behavior of the large resident hogs and the anticipated market value of *T. spiralis*-negative tracer hogs prompted modification of the experimental protocol.

Twelve group 2 tracer hogs were isolated and permitted access to a recently euthanized infected hog carcass for periods of 12, 24, 36 and 48 hours (three hogs per exposure period). Throughout the experiment the tracer hogs were fed their normal grain diet. At the end of the 7-day postexposure period, tracer pigs were returned to lot 2.

e. *Trapping of Rodents and Other Wild Animals*

A variety of traps were used periodically to sample the resident wildlife.

4. Assumptions Inherent in the Methodology

It was assumed that the arbitrary fencing of tracer pigs did not alter the risk of transmission of *T. spiralis* to the hogs, with the exception of those modes specifically prevented. Tracer pigs in lot 2 simulated normal exposure minus cannibalism. Tracer pigs in lot 1 simulated fomite exposure via contaminated boots and feed.

5. Basic Epidemiologic Findings

During the initial visit in January 1984, abundant evidence of hog cannibalism was seen. Hog carcasses or their remains were seen frequently in the lots and shelters. Apparently, long intervals often elapsed between a hog's death and removal of the carcass by farm workers. Newborn pigs farrowed in open lots often were killed and eaten by pigs.

Ten (22%) of 45 muscle specimens from resident hogs were positive for *T. spiralis*. Twenty-eight (20%) of 138 serum samples were serologically positive. Twelve mice, four rats, one raccoon and one opossum were captured. All were free of infection. One cat that was killed by electrocution was found to be infected by muscle digestion.

None of the 39 Group 1 pigs in lot 1 (no exposure to rodents or wild animals) became serologically positive and none were found to be infected at slaughter. Only two of the 33 group 2 pigs that remained in lot 2 throughout the finishing period (exposure to rodents and other wild animals) became serologically positive, but none had *T. spiralis* at necropsy.

Four (25%) of the 12 pigs evaluated in the cannibalism experiment acquired infections within 36 hours after exposure to the infected carcass, and two of four pigs within 12 hours of exposure (Table 13.3). Antibody to *T. spiralis* was found only in pigs with infection confirmed by muscle digestion; however, one pig had only 0.29 LPG and did not develop an appreciable antibody titer by 7 weeks after exposure.

Sixty-six (48.1%) of 137 farm pigs were serologically positive for toxoplasmosis. Over a 5 1/2 month period 26.7% of group 1 pigs and 50% of group 2 pigs seroconverted to *T. gondii* antigen. The eight seropositive group 1 pigs were believed to have been exposed from ingesting grain contaminated with *T. gondii* oocysts, or from oocysts carried on the boots and clothing of the caretakers. The higher rate of seroconversion in group 2 pigs may have been due to ingestion of infected rodents or oocysts from cat feces. Few rodents were seen in the area, however. The experiment clearly demonstrated that *T. gondii* may be transmitted via cannibalism under normal farm conditions. Seven (87.5%) of eight tracer pigs in the cannibalism experiment seroconverted.

The rate of *T. gondii* infection differs from farm to farm. Toxoplasmosis is higher in garbage-fed pigs and in pigs kept on dirt lots than in pigs raised on concrete and fed grain. Garbage feeding attracts cats and rats, thus perpetuating the cat-rodent cycle. Earthworms may also preserve and disseminate *T. gondii* oocysts in dirt lots.

6. Conclusions and Measures Taken

Findings during the 10 years before the present study and data from the present study indicated that transmission had been ongoing on the farm for at least 12 years. The farm partially occupied an earlier garbage dump, which may have provided an initial source of infected rats. Infected rats may have been the original source of *T. spiralis* for the hogs.

Results of the present experiment indicated that cannibalism was a mode of transmission on this farm, and that rodents and/or other wild animals were of little or no importance in the transmission of *T. spiralis*. Before this study, rats may have been an important source of trichinosis on the farm, as earlier studies indicated a high prevalence of infection in rats. By the time of the present study, the farm's rat population had diminished markedly.

Table 13.3 Experimental transmission of *Trichinella spiralis* to hogs via cannibalism

Hog No.	Duration of Exposure (hours)*	Muscle Digestion Results (LPG)			ELISA Results†		
		Diaphragm	Tongue	Tail	Before Exposure	3 Weeks After Exposure	7 Weeks After Exposure
1	12	0.29	0.69	0.11	0.02	0.06	0.11
2	12	0	0	0	0.03	0.09	0.01
3	12	1.60	3.21	0.55	0.10	0.13	0.47
4	24	0	0	0	0.04	0.03	0.04
5	24	0	0	0	0.04	0.08	0.07
6	24	0	0	0	0.06	0.08	0.04
7	36	0	0	0	0.02	0.07	0.04
8	36	1.26	0.20	0.14	0.01	0.16	0.71
9	36	2.94	4.38	0.23	0.07	0.14	0.75
10	48	0	0	0	0.00	0.07	0.03
11	48	0	0	0	0.02	0.10	0.09
12	48	0	0	0	0.00	0.11	0.10

*Tissue from the infected carcass had the following LPG values: diaphragm, 1.13; tongue, 1.02; shoulder, 1.04; rib, 0.70; ham, 0.70; and tail, 0.03.

†ELISA optical density value; positive = 0.25.

LPG = No. of larvae per gram of muscle specimen.

Reprinted with permission from Hanbury, R.D., Doby, P.B., Miller, H.O., and Murrell, K.D. 1986. Trichinosis in a herd of swine: cannibalism as a major mode of transmission. *J.A.V.M.A.* 188:1155-1159.

Tail biting is a second possible mode of transmission. Larvae were recovered from the tails of infected hogs, but the epidemiologic importance of tail biting in porcine trichinosis has not been established convincingly.

Findings in the present study have important implications for the national effort to control trichinosis. One of the control strategies under development is the establishment of slaughterhouse inspection procedures. The notion that the incidence of porcine trichinosis can be reduced by removal of infected pork from the food chain via meat inspection is predicated on the belief that raw garbage is the major source of infection. However, if hog cannibalism or infected rats can maintain the infection (independent of garbage feeding), detection and disposal of infected pork carcasses may not significantly reduce the prevalence of swine trichinosis. Therefore, control strategies should include trichinosis surveillance and traceback procedures and eradication of trichinosis from infected herds. The enactment of the Illinois Trichinosis Control Act on January 1, 1986 provided authority to eliminate trichinosis on this farm. Herd depopulation with indemnity was begun in January 1986.

VI. SUMMARY

Infections may originate (1) iatrogenically, e.g., induced by a clinician's actions, (2) from animal reservoirs or (3) from the environment. Iatrogenic illnesses are those that are induced in a patient by a clinician's actions. Animal reservoirs of disease agents include (1) carrier animals, animals (and human beings) with inapparent infections that are also transmitters (or potential transmitters) of the infectious agent, and (2) intermediate hosts and vectors. Amplifying hosts are 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. Animals that have been exposed to an agent may become carriers. Incubatory carriers are capable of serving as a source of infection while incubating the disease. Convalescent carriers continue to shed infectious organisms after the signs and symptoms of disease have disappeared, i.e., recovery. The environment may be considered a source of infection when the disease agent multiplies there, not requiring any animal host for its continued survival.

Horizontal disease transmission between contemporaries, or animals of more or less the same generation, may occur directly, indirectly or via airborne routes. 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. A second mode of direct transmission is 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 m or less.

Indirect transmission may be vehicle-borne or vector-borne. 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. Vector-borne transmission is generally understood to mean transmission by invertebrate vectors, such as flies, mosquitoes or ticks. It may be 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. Biological transmission requires a period of multiplication, cyclic development or both before the vector can transmit the infective form of the agent. The disease agent may be transmitted vertically (transovarially) between generations of the vector or transstadially from one stage to another within a single generation.

Airborne transmission involves the dissemination of microbial aerosols in the form of 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.

Disease transmission may also occur vertically from animals of one generation to another. 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.

Communicability may be defined as the ease with which a disease agent is spread within a population. One way of expressing communicability is the intrinsic (or basic) reproductive rate (R_0), a dimensionless parameter defined as the average number of secondary cases of infection to which one primary case gives rise throughout its infectious period if introduced into a defined population consisting solely of susceptible individuals. Communicability is affected

by agent, host and environmental factors. Agent factors include the nature of the agent's life cycle and the minimal infective dose. Host factors may appear as heterogeneity in susceptibility to disease due to innate or immune factors. Environmental factors include particle diameter and the microclimate in which the infectious agent finds itself.

Chapter 14

THE COST OF DISEASE

I. 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. Cost-benefit analysis and decision analysis are among the most common (Bennett, 1992). 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.

A. THE "MEASURES OF EFFECT" APPROACH TO ESTIMATING DISEASE IMPACT

The following example takes advantage of the concept of "measures of effect" developed earlier in the text for expressing risk. In this case, risk is expressed in economic terms to determine which risk factors have the greatest economic impact. The history is that of a swine herd experiencing less than optimal performance (see Chapter 11). The history is reproduced along with additional economic factors.

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: A review was made of a year's records and of the relationship of animal performance and management procedures at a swine feedlot in central Kansas (Straw et al, 1985). Aspects of performance that were considered unsatisfactory included (1) slow growth rate of finishing pigs, (2) poor feed conversion, (3) high death rate (especially due to *Haemophilus pneumonia*) and (4) excessive carcass trim at the time pigs were slaughtered. During the year, there was a continuous flow of pigs into and out of the feedlot. Data were used from all groups that had been sold that year.

Analyses were performed on 38 groups containing 9988 pigs. Although overall performance was low, certain groups of pigs (defined as noncases) performed considerably better than others (defined as cases). Comparisons were made between groups in an effort to identify management inputs (risk factors) that could be used to improve overall performance.

Due to the large number of pigs (4400) that could be housed at the feedlot at any one time, certain sound management procedures (all-in/all-out, single source of feeder pigs) could not be implemented. The effects of other management procedures (purchase weight, purchase time,

Table 14.1 Veterinary expenses for pigs entered into a feedlot at two times of the year

Time Pigs Entered the Feedlot	*No. of Groups*	*At Risk (%)*	*Mean Veterinary Expense Per Pig**
Apr to Sep	15	39	$2.92
Oct to Feb	23	61	$4.73
Total	38	Mean	$4.02†

*Total costs of treatment for internal and external parasites, vaccinations, and antibiotics.
†Weighted mean.

Source of data: Straw, B.E., Henry, S.C., and Fleming, S.A. 1985. Interactions of management and animal performance in a swine feedlot. *J.A.V.M.A.* 186:986-988.

Table 14.2 Carcass trim in pigs given various amounts of injectable antibiotics

Mean Amount of Injected Antibiotic Per Pig	*No. of Pigs*	*At Risk (%)*	*Carcass Trim Cost Per Pig*
<4 ml	1249	46	$0.56
>4 ml	1441	54	$3.06
Total	2690	Mean	$1.90*

*Weighted mean.

Source of data: Straw, B.E., Henry, S.C., and Fleming, S.A. 1985. Interactions of management and animal performance in a swine feedlot. *J.A.V.M.A.* 186:986-988.

vaccinations, treatment regimens) on growth rate, feed conversion, death rate and carcass trim were compared among groups of pigs.

Daily death rates (incidences) were calculated by dividing the number of pigs that died on a given day by the total number of pigs present in the lot on the same day. Student's t-test was

Table 14.3 Economic effect of time that pigs entered a feedlot using the measures of effect approach

Simple Risks	
Veterinary costs/pig in exposed*	$4.73
Veterinary costs/pig in unexposed*	$2.92
Veterinary costs/pig overall	$4.02
Prevalence of exposure	61%
Compared Risks	
Relative risk	1.62
Attributable risk/pig	$1.81
Population attributable risk/pig	$1.10
Population attributable fraction	27%

*Exposed pigs entered feedlot October to February; unexposed pigs entered feedlot April to September.
Data from Table 14.1.

Table 14.4 Economic effect of amount of injected antibiotic used upon carcass trim using the measures of effect approach

Simple Risks	
Carcass trim/pig in exposed*	$3.06
Carcass trim/pig in unexposed*	$0.56
Carcass trim/pig overall	$1.90
Prevalence of exposure	54%
Compared Risks	
Relative risk	5.46
Attributable risk/pig	$2.50
Population attributable risk/pig	$1.34
Population attributable fraction	71%

*Exposed pigs injected with >4 ml antibiotic; unexposed pigs injected with <4 ml antibiotic.
Data from Table 14.2.

used to compare performance between groups of pigs. The Chi-square test was used to compare mortality rates.

The factor having the greatest influence on performance was the month of entry of pigs into the feedlot. Pigs that entered the feedlot between April and September performed better than did pigs entering between October and February (Table 14.1). The amount of carcass trim per pig was significantly less ($P<0.001$) in pigs that were treated with less than 4 ml of antibiotic by injection versus pigs that were treated with more than 4 ml of antibiotic ($0.56 versus $3.06 per pig slaughtered, respectively; Table 14.2).

Single source pigs did not perform better than multiple source pigs, nor did heavier pigs versus lighter pigs. However, total veterinary costs per pig were lower for pigs that weighed more than 27 kg on entry into the feedlot than for pigs that weighed 27 kg or less ($3.53 versus $4.70, respectively; $P<0.01$). The average daily death rate among pigs that failed to reach market weight within 150 days after entry into the feedlot (0.0104) was nearly twice that of the pigs that reached market weight before 150 days (0.0054) (see Figure 5.3).

The investigators recommended that the producer (1) start pigs only during spring and summer months, (2) use oral antibiotic therapy if possible to avoid carcass trim at slaughter, (3) market all animals by 150 days after entry into the feedlot (regardless of age) and (4) use a *Haemophilus* vaccine of proven efficacy. However, the actual economic benefit of adopting these recommendations was not estimated.

An analysis of veterinary costs and carcass trim based on presence or absence of risk factors (time of entry into feedlot and volume of antibiotic used, respectively) appears in Tables 14.3 and 14.4. Cost figures are drawn from Tables 14.1 and 14.2, respectively. The prevalence of exposures for Table 14.3 are calculated from group data, whereas for Table 14.4 it is based on actual pig numbers. From Table 14.3 it can be seen that by starting pigs during spring and summer months, veterinary costs per pig can be reduced $1.10, or 27% (population attributable fraction). The analysis in Table 14.4 shows that by reducing the amount of injected antibiotic below 4 ml, carcass trim per pig can be reduced $1.34, or 71%.

A similar analysis could be performed on feed efficiency based on the cost per pound of gain. This sort of analysis gives the veterinarian or producer a better idea of where to start first in reducing economic losses due to endemic disease. The measures of effect approach does not, however, include the cost of the disease control program in its analysis.

B. PARTIAL BUDGETING AND BENEFIT-COST ANALYSIS

1. Partial Budgeting

In order to estimate benefits and costs to producers of a specific disease control program, partial budget analysis rather than total budget analysis is frequently used. 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 property taxes) are ignored. Determining costs specific to a single disease outbreak requires partial budget analysis.

Partial budgeting usually places farm budget items into one of four categories (Martin et al, 1987):

(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.

2. Benefit-Cost Analysis

Benefit-cost analysis is a method for calculating a benefit-cost ratio. Dollar values can be ascribed to the budget categories described previously which are then assigned to one of two broad categories: benefits (B) and costs (C). The ratio of benefits to costs (B/C) is the benefit-

cost ratio and is an index of the dollar value of benefits that can be expected from a given cost investment.

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 ratio is less familiar to most veterinarians. 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. Costs for a disease eradication program accrue during the relatively short life of the program. Benefits after disease eradication 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 B/C ratio would not be $1000/$500, or 2. This is because $1000 invested today will be worth considerably more than $1000 5 years from now. If, for example, we assume a 10% interest rate over the next 5 years, $1000 5 years from now would be equivalent to $620.90 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 present value is

$$PV = \frac{1}{(1 + r)^n} \times FV$$

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 time in years. As the time (n) before a benefit accrues increases, the present value of future benefits decreases.

II. DECISION ANALYSIS

In most cases in veterinary practice, the prognosis or economic impact of medical decisions is not certain. The best option, e.g., 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. Models of the decision-making process provide a graphic method to aid the decision maker.

A. STEPS IN BUILDING A DECISION TREE

Decision analysis is a process for analyzing complex choices by the use of decision trees (Pauker and Kassirer, 1987; Kassirer et al, 1987; Smith, 1993). There are 3 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.

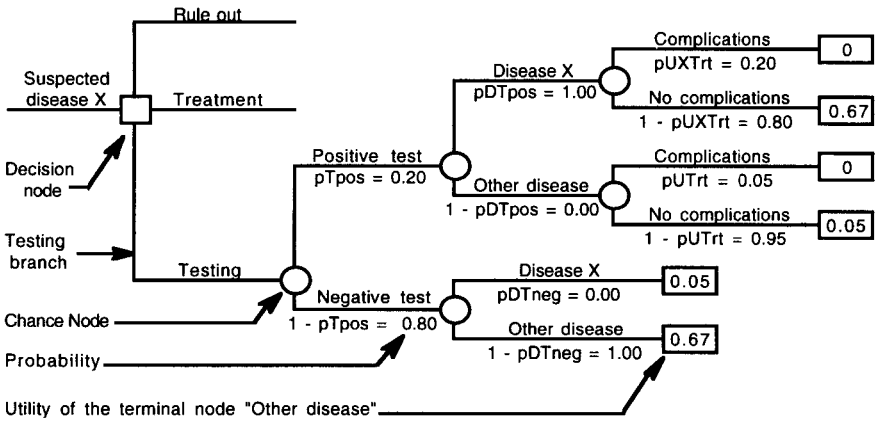


Figure 14.1 Diagram of a portion of a decision tree for "Disease X" to illustrate basic concepts of decision analysis.

The decision node is designated by a square, chance nodes by circles, and terminal nodes are represented 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 * 0.05) + (1.0 * 0.67) = 0.67. By risk analysis, the probability of death (utility = 0) for the testing branch of this decision tree is (0.2 * 1.0 * 0.2) + (0.2 * 0.0 * 0.05) = 0.04. (From Smith, R.D. 1993. Decision analysis in the evaluation of diagnostic tests. *J.A.V.M.A.* 203:1184-1192. With permission.)

B. 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 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. A *terminal node* represents a final outcome with no further significant options or consequences.

C. 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.

D. 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.

E. 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.

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.

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 (Madison et al, 1984; Fetrow et al, 1985). 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.

3. Risk Profile Analysis

Fold back of the decision tree does not express 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.

III. 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 economic calculations involve use of the difference between decreased revenues received and decreased costs incurred (Hoblet et al, 1987).

A. 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. In summary, disease prevention focuses on elimination of risk factors.

Returning to the preceding example of performance of swine in a Kansas feedlot, the time of entry into the feedlot, dose of injected antibiotic, duration of stay in the feedlot and failure to vaccinate were all *risk factors* associated with production and economic losses.

B. 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 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 economic criteria and the Food Safety and Inspection Service who inspects the producer's livestock must consider public health risks of the disease.

C. 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. Returning to the case report of trichinosis on an Illinois swine farm, epidemiologic investigations indicated that swine were the only reservoir of the infection for other swine. Garbage feeding was not practiced. Therefore, the approach used was to eradicate the disease from the farm by depopulating the herd and replacement with trichinosis-free stock. In theory, trichinosis should not appear on these premises again.

Twelve major livestock diseases and pests have been eradicated from the United States since 1884.

Twelve major livestock diseases and pests have been eradicated from the United States since 1884. These are contagious bovine pleuropneumonia, Texas cattle fever, foot and mouth disease, dourine, glanders, fowl plague, vesicular exanthema, screwworms, sheep scabies, Venezuelan equine encephalitis, exotic Newcastle disease and hog cholera. 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.

Very high levels of artificially induced herd immunity are required to eradicate diseases whose intrinsic reproduction rates are high (see Table 13.1). The relatively small value of R_0 for smallpox, and corresponding low level of herd immunity that must be artificially induced, may partly explain the success of the global smallpox eradication campaign. Other factors are the obviousness of the disease and availability of an effective vaccine. In contrast, the high values of R_0 for malaria 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.

IV. CASE STUDIES

A. ECONOMIC ASSESSMENT OF A PSEUDORABIES EPIZOOTIC, BREEDING HERD REMOVAL/REPOPULATION AND DOWNTIME IN A COMMERCIAL SWINE HERD (HOBLET ET AL, 1987)

Partial budget analysis is used to break down the costs and benefits of a disease eradication program.

1. Introduction

Pseudorabies is a disease of swine caused by the pseudorabies virus (PRV), a herpesvirus. Initial production losses after the introduction of PRV onto a swine farm can be severe and include increased suckling pig mortality, increased frequency of stillbirths, fetal mummification, abortions and subsequent sow infertility. However, after the initial pseudorabies epidemic (epizootic or outbreak) in a farrowing operation, documented observations of specific production losses and associated economic consequences are limited (Hoblet et al, 1987).

Currently, there is considerable interest in methods for pseudorabies eradication or control, and several pilot projects involving various procedures are in progress. Three basic plans for elimination of PRV from a swine herd are generally recognized. These include (1) test and removal of animals that test positive (with or without vaccination of breeding stock), which is intended to rebuild a herd from seronegative adults; (2) offspring segregation, which is intended to develop a herd using the offspring of seropositive adults; and (3) depopulation and repopulation (Anonymous, 1987). Any one of these strategies may include vaccination.

An outbreak of pseudorabies was recognized on a 150-sow farrow-to-finish swine operation in north central Ohio on March 3, 1983. Pigs on this farm were believed to have been exposed to PRV through spread of the virus from a neighboring feedlot. The chronology of the pseudorabies epidemic in the farrowing unit is depicted in Figure 14.2. Clinical signs included vomiting, pyrexia, occasional ataxia and increased preweaning mortality of 5- to 7-day-old pigs. Mean preweaning mortality was 16.8% before the epidemic, 37.7% during the epidemic (February and March 1983) and 19.1% after the epidemic. By April 13, 98.5% of pigs tested had developed detectable antibody titers, but clinical signs of pseudorabies had abated.

Results of a serologic surveillance program revealed that pigs born subsequent to the epidemic were serologically negative for PRV (after loss of colostral antibody), indicating that PRV transmission was not occurring during summer and autumn of 1983. Therefore, a depopulation/repopulation plan was agreed on between the producer and the Ohio Department of Agriculture to sell all PRV-positive animals and to begin farrowing again in spring 1984 with the seronegative young stock.

2. Purpose of the Study

The purpose of this study was to measure the production and economic impacts that could be attributed to a pseudorabies epidemic, breeding herd removal/repopulation and downtime in the swine operation.

3. Epidemiologic Methodology

a. *Source of Data*

Farm production data recorded before the epidemic of pseudorabies were used to establish baseline values for suckling pig mortality, litter size, stillbirths and sow culling. Records were available on the number of sows farrowed, the number of pigs born alive and the number of pigs that died before weaning from February 1982 (1 year prior to the epidemic) through November 1983.

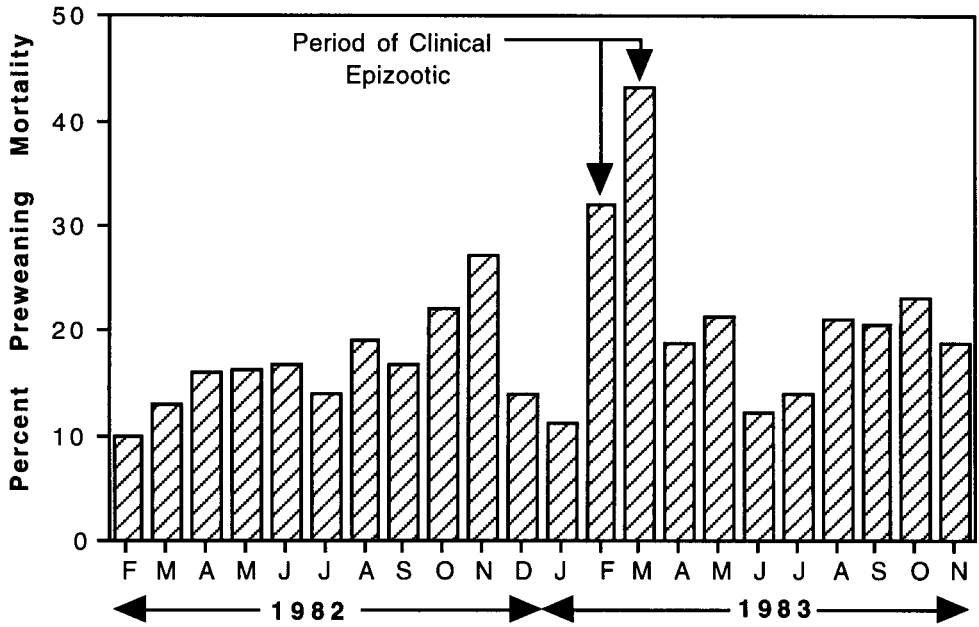


Figure 14.2 Suckling pig mortality before weaning. A clinical epidemic (epizootic) of pseudorabies occurred among pigs farrowed during February and March, 1983. Mean preweaning mortality was 16.8% before the epidemic, 37.7% during the epidemic and 19.1% after the epidemic. (Reprinted with permission from Hoblet, K.H., Miller, G.Y., and Bartter, N.G. 1987. Economic assessment of a pseudorabies epizootic, breeding herd removal/repopulation, and downtime in a commercial swine herd. *J.A.V.M.A.* 190:405-409.)

b. *Calculation of Economic Impact*

Partial budgeting (which focuses on discrete input-output relationships) was used to calculate the economic impact associated with the pseudorabies epidemic. The major causes of loss used in estimating the economic impact were (1) changes in mortality of suckling pigs, (2) changes in numbers of stillbirths, (3) changes in percentages of sows culled and (4) forced sale/repopulation.

Suckling Pig Mortality – Economic losses associated with suckling pig mortality were calculated by first estimating lost revenues and then adjusting for decreased variable production costs not incurred because of death loss. The increased number (greater than baseline) of unweaned (younger than 4 weeks old) pig deaths was multiplied by the mean weight for market pigs on the farm (2.39 hundredweight [cwt]) times the mean price received for market pigs in 1983 ($47.35/cwt) minus the estimated mean variable costs of production ($37.21/cwt), or $10.14. The general formula for dollar loss was

(increased deaths) x (market pig weight) x ($10.14)

Stillborn Pigs – The case definition for a stillbirth was any fully developed fetus found dead on passage through the birth canal. Economic loss from increased numbers of stillborn pigs was treated in the same way as suckling pig mortality. The general formula for dollar loss was

(increased stillborns) x (market pig weight) x ($10.14)

Culled Sows – Economic losses from excessive culling during the epidemic were estimated by multiplying the increased numbers of sows culled (post-epidemic cull rate minus normal cull rate) times duration of the epidemic before depopulation (in months) times replacement costs minus slaughter value ($50). The general formula for dollar loss was

(excess cull rate) x (months) x ($50)

Forced Sale of Sows – The number of sows sold to effect eradication was estimated by subtracting the normal culling rate times 7 months (downtime) from the normal mean farm sow population. This number was multiplied by the replacement cost minus slaughter value ($50) to estimate costs for the forced sale of sows. The general formula for dollar loss was

(mean sow population - 7 x [normal culling rate]) x ($50)

Forced Sale of Boars – Dollar loss for forced sale of boars was calculated as in the previous example except that $200 was used rather than $50 as the replacement cost minus slaughter value. The general formula for dollar loss was

(mean boar population - 7 x [normal culling rate]) x ($200)

Downtime (No Farrowing) – Since no farrowing occurred over the 7-month period from December 1983 to July 1984, lost revenue had to be estimated. This was estimated by multiplying the estimated number of pigs normally marketed per month (95% of pigs weaned) times 7 months times the normal market weight (2.39 cwt) times the difference between the market price for pigs and the variable costs of production, or $10.14. The general formula for dollar loss was

(pigs sold per month) x 7 x (market weight) x ($10.14)

The Z test was used to test for significant differences in suckling pig mortality, stillbirths and sow culling rates when pre-epidemic and post-epidemic periods were compared. The t-test was used to compare mean litter size before and after the pseudorabies epidemic. Differences were considered significant if $P < 0.05$.

4. Assumptions Inherent in the Methodology

The assumption was made that any increase in suckling pig mortality, stillborns, or sow culling rate over normal baseline levels was due to pseudorabies. For purposes of calculation of losses from increased suckling pig mortality and stillborn pigs, it was assumed that 95% of pigs weaned would be finished to 2.39 cwt on the farm. It was further assumed that differences in costs of production between a stillborn pig and a pig the age of those dying of pseudorabies were minimal.

For long-term economic analysis of the eradication strategy it was assumed that (1) post-epidemic losses were entirely the result of pseudorabies, (2) that revenues and costs of production remained fixed and similar to those experienced during the study period and (3) that without farm eradication, such losses would continue indefinitely at the same rate. Furthermore, a 10% interest rate was assumed.

5. Basic Epidemiologic Findings

Testing from January to March 1984 indicated that all swine on the farm (828 pigs tested) were seronegative for pseudorabies. Quarantine was lifted in March 1984. However, testing

Table 14.5 Partial budget of benefits and costs of a pseudorabies control program for a 150-sow farrow-to-finish operation in Ohio

Nature of Benefit or Loss	*Benefits ($)*	*Costs ($)*
1. Additional returns (due to adoption of the proposed control program)		
Reduced suckling pig mortality		
During epidemic (2 mo)	2,234	
After epidemic (8 mo)	944	
Reduced stillbirths		
During epidemic (2 mo)	1,059	
After epidemic (8 mo)	1,105	
Reduced sow culling	623	
2. Foregone returns (reduced numbers of culled animals)*		0
3. Additional costs incurred (due to the control program – drugs, management procedures, etc.)		
Breeding herd removal/repopulation		
Sows		6,415
Boars		1,140
Downtime		34,655
4. Costs no longer incurred (if control program is implemented – salvage treatments)	0	
Total benefits (1 + 4)	$5,965	
Total costs (2 + 3)		$42,210

*"Foregone returns" incorporated into "Reduced sow culling" above as replacement cost minus market value.

Source of data: Hoblet, K.H., Miller, G.Y., and Bartter, N.G. 1987. Economic assessment of a pseudorabies epizootic, breeding herd removal/repopulation, and downtime in a commercial swine herd. *J.A.V.M.A.* 190:405-409.

after the July 1984 farrowing revealed that a previously seronegative gilt had seroconverted. Subsequent to this finding and primarily because of problems with the lending agency, the owner depopulated. A partial budget analysis of production losses is summarized in Table 14.5.

6. Conclusions and Measures Taken

Clearly, losses associated with removal/repopulation of the breeding herd and subsequent downtime greatly exceeded the potential benefits from the disease control strategy. Assuming

that post-epidemic losses for the period monitored of $2672 ($944 + $1105 + $623) were to continue indefinitely, the authors estimated that the disease could be endemic for 22 years before accumulated costs would exceed the costs of the removal/repopulation strategy, using standard discounting techniques. If a clinical epidemic were assumed to occur on alternate years, 12 years would pass before the cost of disease would exceed the costs of the eradication program.

From an economic standpoint it would appear that the method of pseudorabies eradication used on this farm was less than ideal. The costs associated with downtime far exceeded the costs of the disease, and should be limited as much as possible in any eradication scheme. It is hoped that future research efforts and pseudorabies pilot eradication/control projects will address economic costs of alternative plans for pseudorabies eradication.

B. DECISION ANALYSIS OF A HEARTWORM DIAGNOSTIC TEST (SMITH, 1993).

Adapted with permission from the Journal of the American Veterinary Medical Association.

Decision analysis is used to evaluate the clinical usefulness of a serodiagnostic test.

1. Introduction

Canine heartworm disease is a potentially severe but treatable disease with signs that mimic those of scores of other diseases. Infections may be microfilaremic or amicrofilaremic (occult), and occult infections are frequently diagnosed with the aid of serologic tests that detect circulating antigens of *Dirofilaria immitis*. Although commercially available antigen detection kits are good, none is 100% sensitive and specific. Thus the possibility of false positive or false negative test results exists, with important penalties for both.

Although properties such as test sensitivity and specificity are useful for comparing performance characteristics of tests, they provide little information about the usefulness of a test in clinical practice. Other factors, such as disease prevalence (or likelihood) and the relative desirability and likelihood of the various possible outcomes, affect how test results are used to choose between management options.

2. Purpose of the Study

The purpose of this study was to determine under what conditions heartworm antigen detection test results can be accepted or ignored when managing a suspected case of occult canine heartworm disease.

3. Epidemiologic Methodology

a. *Decision Tree Construction*

A heartworm decision tree (Figure 14.3) was constructed with one decision node and branches representing three distinct management options. The *rule out branch* represents the decision to eliminate other differential diagnoses before considering heartworm disease, even if an antigen detection test is performed and the result is positive. This option might be chosen if the likelihood of heartworm disease is considered to be low. The *treatment branch* represents the decision to treat the dog as if it has heartworm disease, even if an antigen detection test is performed and the result is negative. This alternative might be chosen when the likelihood of heartworm disease is considered to be high.

The *testing branch* of the tree represents the decision to let the results of the antigen detection test guide subsequent case management. For this branch, the pretest probability of heartworm disease, test sensitivity, and test specificity were combined in various formulas (see below) to calculate the probability of a positive test result, the probability of heartworm disease given a positive test result (predictive value of a positive test), and the probability that

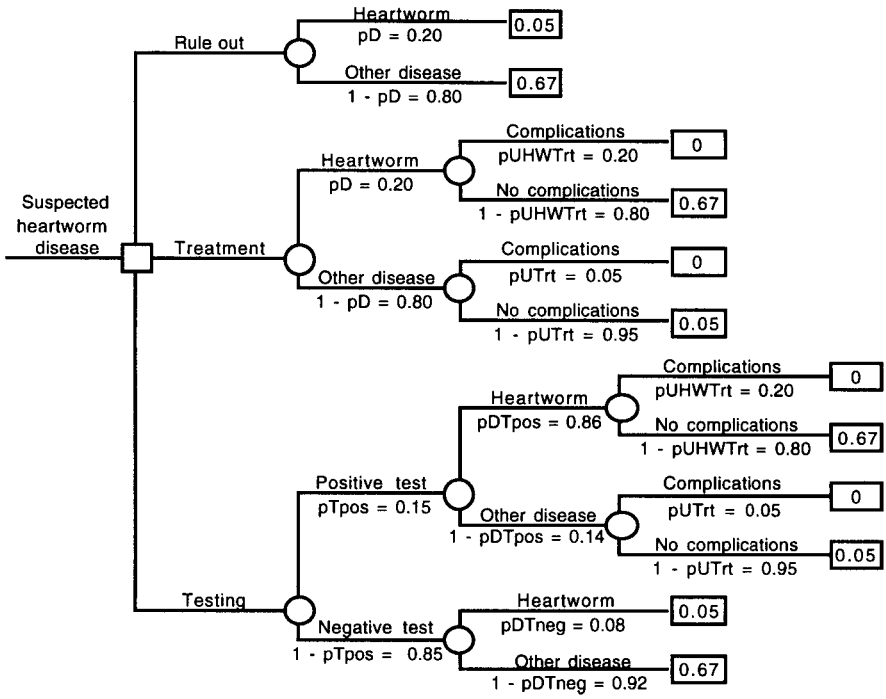


Figure 14.3 Decision tree for clinical management of a dog suspected of having occult heartworm disease. The decision tree is shown with baseline probability and utility values.

The probability of a positive test result (pTpos) = (sensitivity * pD) + [(1 - specificity) * (1 - pD)], where pD is the prior probability of heartworm disease. The predictive value of a positive test result (probability of disease given a positive test result, pDTpos) = (sensitivity * pD) ÷ pTpos. Similarly, the probability of disease given a negative test result (pDTneg) = [(1 - sensitivity) * pD] ÷ (1 - pTpos). The predictive value of a negative test result (probability of not having the disease given a negative test result) = 1 - pDTneg. pUHWTrt = probability of complications from thiacetarsamide treatment of heartworm disease leading to death. pUTrt = probability of complications from thiacetarsamide treatment of non-heartworm disease leading to death. (From Smith, R.D. 1993. Decision analysis in the evaluation of diagnostic tests. *J.A.V.M.A.* 203:1184-1192. With permission.)

the dog does not have heartworm disease given a negative test result (predictive value of a negative test). Chance nodes include values or expressions for the likelihood of disease, likelihood of a particular test result, predictive value of the diagnostic test, and likelihood of a favorable or unfavorable outcome.

b. *Assignment of Variables*

The a priori likelihood that a particular animal will have the disease (prior probability) depends on the prevalence of disease in the area or clinical setting, the characteristics of the animal, and the results of other tests which may previously have been performed. The utility of each final outcome (terminal node) was expressed as prognosis for complete recovery. Utility values were derived from studies on the clinical course of dogs with advanced heartworm disease. With the exception of death, the utility of which is 0, all other probabilities and utilities are variables, the values of which can be changed for the purpose of sensitivity analysis. Patient death was depicted explicitly in the decision tree to facilitate risk profile analysis. All variables were assigned a baseline value for initial fold back of the tree (Table 14.6).

Table 14.6 Baseline probabilities and utilities used in heartworm decision tree (Figure 14.3)

Event	*Notation*	*Probability*	*Utility*
Presence of heartworm infection	pD	0.20	- - -
Likelihood of a positive test result			
In heartworm infection (sensitivity)	pT+/D+	0.650	- - -
In non-heartworm disease (false positive rate)	pT+/D-	0.027	- - -
Likelihood of a negative test result			
In heartworm infection (false negative rate)	pT-/D+	0.350	- - -
In non-heartworm disease (specificity)	pT-/D-	0.973	- - -
Complications from Caparsolate therapy			
In heartworm-infected dogs	pUHWTrt	0.20	- - -
In non-heartworm disease	pUTrt	0.05	- - -
Recovery without severe sequelae:			
In Caparsolate-treated heartworm infection	- - -	- - -	0.67
In untreated heartworm infection	- - -	- - -	0.05
Specific therapy of non-heartworm disease	- - -	- - -	0.67
Untreated non-heartworm disease	- - -	- - -	0.05

c. *Probability of a Positive Test Result*

The probability of a positive test result (pTpos) = (sensitivity * pD) + [(1 - specificity) * (1 - pD)], where pD is the prior probability of heartworm disease.

d. *Predictive Value of a Positive Test*

The predictive value of a positive test result (probability of disease given a positive test result, pDTpos) = (sensitivity * pD) ÷ pTpos.

e. *Predictive Value of a Negative Test*

The probability of disease given a negative test result (pDTneg) = [(1 - sensitivity) * pD] ÷ (1 - pTpos). Therefore, the predictive value of a negative test result (probability of not having the disease given a negative test result) = 1 - pDTneg.

f. *Utilities*

The utility of untreated heartworm disease was given a baseline value of 0.05, reflecting the low likelihood that dogs with heartworm disease will improve when given only supportive therapy (restricted activity, low-sodium diet, and furosemide). The baseline prognosis for recovery (utility) after diagnosis and treatment of other diseases on the differential list was arbitrarily set at 0.67, equal to that assigned to the baseline prognosis for recovery from heartworm disease.

For dogs suffering from heartworm disease, the probability that complications from thiacetarsamide treatment will lead to death (utility = 0) is no more than 0.20. Of the dogs that do not develop complications, 67% will become free of clinical signs, and 33% will continue to have clinical signs. For dogs that are not suffering from heartworm disease, there is still a chance (≤5%) of death associated with thiacetarsamide administration. The baseline prognosis

for recovery, without specific treatment, from diseases other than heartworm disease was arbitrarily given a value of 0.05, the same as the utility for untreated heartworm disease.

g. *Analyses Performed*

Fold back and risk profile analysis were performed, using baseline values. In addition, because prior probability will affect predictive values of positive and negative test results, one-way sensitivity analysis was used to determine the range of prior probability values over which reliance on test results is the preferred management option, e.g., the testing band (see Chapter 4). Since test sensitivity varies with worm burden (Courtney et al, 1990), the combined effect of prior probability and test sensitivity upon the testing band was evaluated by use of two-way sensitivity analysis for test sensitivities ranging from 50 to 100%, corresponding to increasing worm burdens. Specificity was fixed at 97.3%.

Another factor that can influence the degree of reliance on test results is the penalty for misdiagnosis. This was explored through three-way sensitivity analysis of the effect of the relative prognosis of heartworm disease versus that for other differential diseases on the testing threshold. The prognoses for dogs with heartworm disease that were treated and for dogs with other diseases that were appropriately diagnosed and treated were varied independently at five levels of prior probability of heartworm disease ranging from 5 to 45%. All other variables were held at baseline levels.

4. Assumptions Inherent in the Methodology

It was assumed that any biases introduced by the choice of baseline values would be compensated for through one-, two-, and three-way sensitivity analysis. The utility assigned to each terminal node in the heartworm decision tree was expressed as a prognosis as there was no way to represent the value that people place on their pets. Utility values were derived from studies on the clinical course of dogs with advanced heartworm disease, thus representing a "worst case scenario." The baseline prognoses (utilities) for recovery from heartworm and non-heartworm disease were made equivalent so as not to bias the interpretation of other variables that affect the decision to rely on heartworm test results.

5. Basic Epidemiologic Findings

Fold back of the heartworm decision tree, using baseline values (Figure 14.3; Table 14.7) revealed that reliance on the diagnostic test provided the highest expected utility (expected prognosis: 0.60), followed by rule out of other diseases (0.55), and empiric heartworm treatment (0.15). Although fold back suggested that there was little reason for choosing the testing option over the rule outs option, a risk profile analysis (Table 14.7) revealed that the chance of an unfavorable outcome (expected utility ≤ 0.05) was 12% for the testing branch versus 20% for the rule out branch. Thus, letting diagnostic test results guide patient management is clearly the best choice.

Under baseline conditions, the testing band (Figure 14.4) ranged from a prior probability of 3% (testing threshold) to a prior probability of 78% (treatment threshold). Thus, when the prior probability of heartworm disease is $< 3\%$, then rule out of other diseases is the preferred option, even if test results are positive. If it is $> 78\%$, then heartworm treatment should be instituted, even if test results are negative. Test results should guide treatment decisions in all other cases. The recommendation to eliminate other differential diagnoses when the likelihood of heartworm disease is low does not mean that the use of antigen tests to screen dogs for heartworm disease should be discouraged. Under these conditions, the utility of treating or not treating diseases other than heartworm is 1.00, the prognosis that a healthy, uninfected dog will remain healthy. Repeating the testing band analysis, using this value, the testing threshold becomes 0.4%. As long as the prior probability for heartworm disease is $> 0.4\%$, as is the case in most heartworm-endemic regions, testing for heartworm disease is the best option.

Table 14.7 Comparison of fold back and risk profile analysis of the heartworm decision tree (Figure 14.3)

Management option	*Fold back analysis (average expected utility)*[a]	*Risk profile analysis*[b]		
		*Expected utility = 0.00**	*Expected utility = 0.05*†	*Expected utility = 0.67*§
Pursue other diagnoses	0.55	0%	20%	80%
Treatment	0.15	8%	76%	16%
Testing	0.60	3%	9%	88%

[a]Average prognosis for each management option where that option is chosen for a large number of animals.

[b]Likelihood of a favorable (expected utility, 0.67) or unfavorable (expected utility, 0.0 or 0.05) outcome.

*Represents death because of thiacetarsamide administration to dogs with or without heartworm disease.

†Represents prognosis for recovery without treatment of dogs with or without heartworm disease.

§Represents prognosis for recovery when dogs with heartworm disease or with some other disease are given appropriate treatment.

From Smith, R.D. 1993. Decision analysis in the evaluation of diagnostic tests. *J.A.V.M.A.* 203:1184-1192. With permission.

Higher worm burdens are associated with an increase in test sensitivity and a decrease in the percentage of false negative test results (Courtney et al, 1990). Two-way sensitivity analysis was used to evaluate the interaction between test sensitivity and prior probability of heartworm disease (Figure 14.5). As test sensitivity increased, the treatment threshold increased from 72% (1-2 worms; sensitivity, 53.2%) to 100% (>20 worms; sensitivity, 100%). Worm burden has no effect on test specificity or the likelihood of false positive test results. Because test specificity was high and remained unchanged, the testing threshold only increased from 4% to 6% as test sensitivity increased. Results indicate that one can have a high degree of confidence in heartworm test results, especially if more than a few worms are present.

The penalty for misdiagnosis was evaluated (Figure 14.6) by constructing indifference curves for various prior probabilities of heartworm disease. For a given prior probability, any point below the indifference curve would mean that diagnostic test results should guide patient management; whereas any point above the indifference curve would mean that other differential diagnoses should be pursued, regardless of test result. In most cases, the prognosis for treatment of other differential diseases must exceed that for treatment of heartworm disease before a positive heartworm test result can be ignored, even when the likelihood of heartworm disease is low (prior probability = 5%). The high clinical usefulness of testing is, in part, a result of the relatively high sensitivity and specificity of currently available heartworm antigen tests.

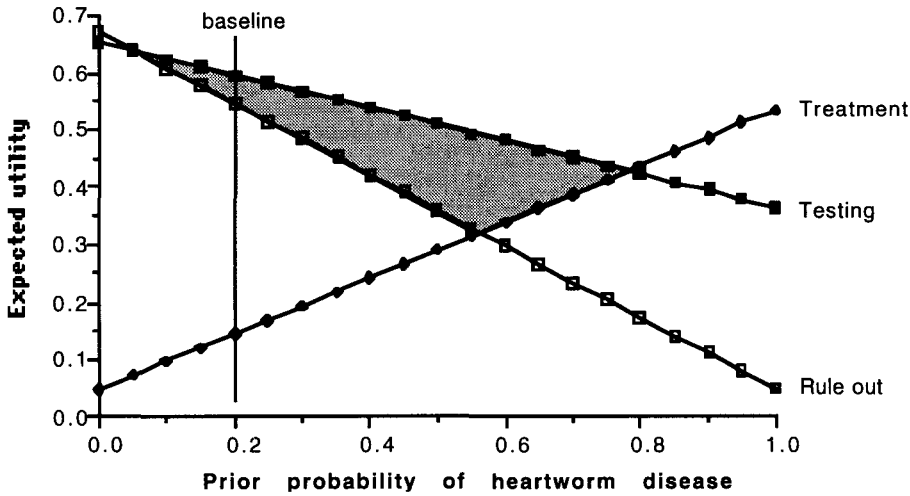


Figure 14.4 One-way sensitivity analysis of the heartworm decision tree to determine the effect of prior probability of heartworm disease on expected utility.

Analyses were performed for prior probabilities ranging from 0 to 100%. Treatment represents expected utility for empiric treatment; Testing represents expected utility for basing treatment on the results of serologic testing for heartworm disease; Rule out represents the expected utility of attempting to eliminate other differential diagnoses. The shaded area represents the testing band, the range of prior probabilities for which testing for heartworm disease is the best management option. (From Smith, R.D. 1993. Decision analysis in the evaluation of diagnostic tests. *J.A.V.M.A.* 203:1184-1192. With permission.)

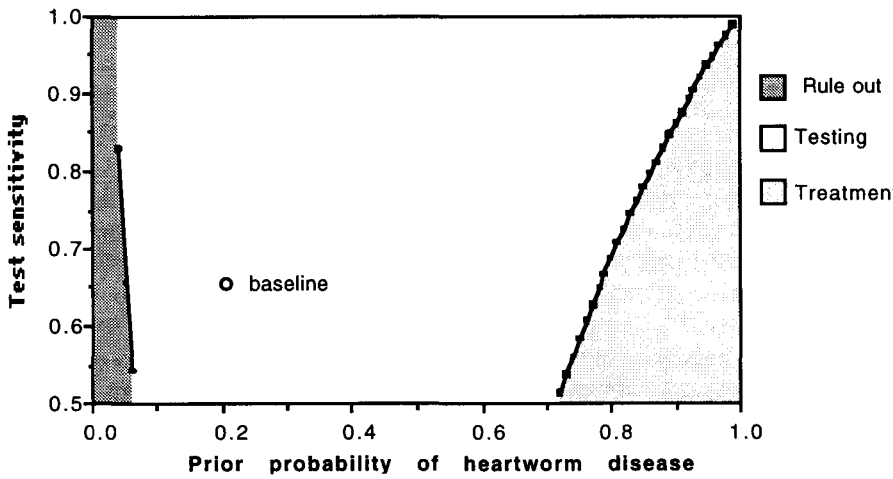


Figure 14.5 Two-way sensitivity analysis of the heartworm decision tree to determine the effect of test sensitivity and prior probability of heartworm disease on the testing and treatment thresholds.

Analyses were performed for test sensitivities ranging from 50 to 100%, corresponding to increasing worm burden. Testing threshold represents the prior probability at which pursuing other differential diagnoses regardless of test results and letting test results dictate treatment are equally good management options. Treatment threshold represents the prior probability at which letting test results dictate treatment and treating regardless of test results are equally good management options. Specificity was held at 97.3%. (From Smith, R.D. 1993. Decision analysis in the evaluation of diagnostic tests. *J.A.V.M.A.* 203:1184-1192. With permission.)

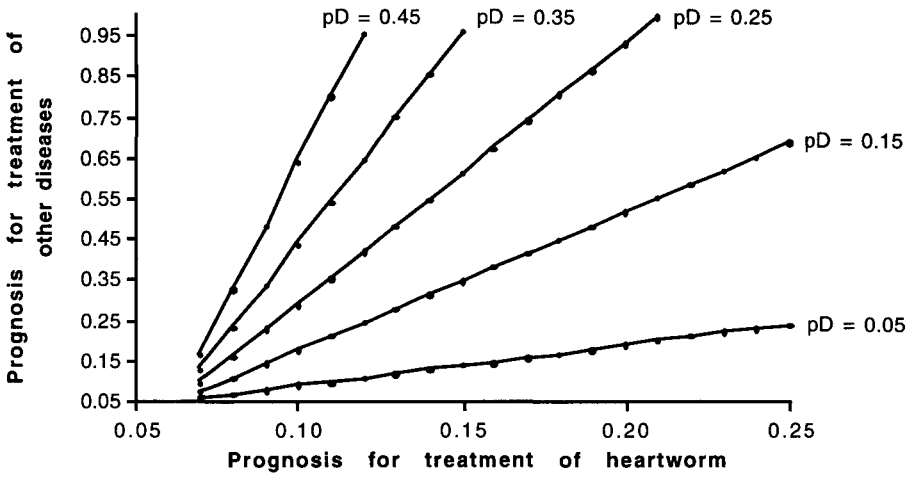


Figure 14.6 Three-way sensitivity analysis of the heartworm decision tree to determine the effect that prior probability of heartworm disease (pD), prognosis for treatment of heartworm disease, and prognosis for treatment of other diseases would have on expected utility.

Three-way sensitivity analysis results in a series of indifference curves. For a given prior probability of heartworm disease, any point above the curve would indicate that other differential diagnoses should be pursued, whereas any point below the curve would indicate that results of testing should dictate treatment. (From Smith, R.D. 1993. Decision analysis in the evaluation of diagnostic tests. *J.A.V.M.A.* 203:1184-1192. With permission.)

6. Conclusions and Measures Taken

The width of the testing band and the values for testing and treatment thresholds suggest that heartworm antigen testing can be used with confidence in a variety of clinical settings ranging from screening of clinically normal dogs to differential diagnosis of heartworm disease in dogs with severe clinical signs. The prognosis for treatment of dogs with heartworm disease included data from dogs in advanced stages of heartworm disease and, therefore, probably represented a worst-case scenario. In general, as the prognosis for heartworm therapy increases, both the testing and treatment thresholds (prior probability at which either testing or empiric therapy is the best choice) decrease. It was difficult to find a set of conditions under which test results should be ignored.

V. SUMMARY

A producer's decision as to whether to institute any sort of disease control program will be based, in large part, on economic considerations. In order to better target a disease control program, some sort of economic analysis is usually necessary. The "measures of effect" approach is similar to that used to express risk, with the exception that cost figures are substituted for incidence rates. This sort of analysis gives the veterinarian and producer a better idea of where to start first in reducing economic losses due to endemic disease. The measures of effect approach does not, however, include the cost of the disease control program in its analysis.

In order to estimate benefits and costs to producers of a specific disease control program, partial budget analysis is frequently used. The part of the enterprise budget affected by the dis-

ease is separated out from unrelated costs. Fixed costs (such as property taxes) are ignored.

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 and (4) *costs no longer incurred* such as veterinary expenses.

Benefit-cost analysis is a method for calculating a benefit-cost ratio. Partial budget categories are assigned to one of two broad categories: benefits and costs. The ratio of benefits to costs suggests the dollar value of benefits that can be expected from a given cost investment.

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. Using a discount rate, disease control program benefits and costs that accrue in the future are discounted to present values.

In most cases in veterinary practice, the prognosis or economic impact of medical decisions is not certain. The best option, e.g., 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. In decision tree analysis a decision tree is constructed incorporating decision nodes (points in the model where a decision must be made) and chance nodes (points where the branch of the tree taken is governed by a chance event with its associated probability). The probability of a positive or negative test result is estimated from the sensitivity, specificity and prevalence of the disease. Bayes' theorem is used to estimate the posterior probability of disease given a positive test result. After the tree is constructed, the expected utility of a particular decision may be computed by weighting the value (in terms of prognosis or dollars) of each potential outcome with the probability that the outcome will occur, given a particular decision, and then summing the weighted values for all potential outcomes for a given branch of the decision tree. This is referred to as "folding back" the tree. The branch with the highest expected utility will maximize favorable outcomes over a series of such choices.

Since a fold back does not give the distribution of how likely each result is, the best choice may not be the branch with the highest expected utility. In this case a risk profile can be used to estimate the risk of unfavorable outcomes for each branch on the decision tree.

One of the principal benefits of decision tree analysis is that sensitivity of a choice to its underlying assumptions can be tested. At the breakeven point, the expected values for two or more interventions are equal. In this case, the decision about which approach to use can be made on grounds other than the prognosis.

Ultimately the practitioner must devise a plan for the reduction of disease in the population. This may be accomplished through disease prevention, control (including treatment) or eradication. The objective of disease prevention is to forestall disease transmission or the occurrence of clinical signs. Disease control is aimed at reducing the frequency of disease to a tolerable level. Eradication is the complete elimination of a disease agent from the environment. The feasibility of eradication depends on meeting one or more of the following conditions: (1) an effective means for identification of reservoirs (carriers), (2) an effective method for destruction of the agent in reservoirs (or the reservoirs themselves), (3) small host range (preferably a single host) and (4) single or limited spectrum of disseminating mechanisms that can be readily manipulated.

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.

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.

Apparent prevalence The prevalence of disease estimated on the basis of diagnostic tests (compare with real 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.

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).

Bias A mental leaning or inclination. Not leaving the mind indifferent. Syn. – tendency, inclination, propensity, disposition, bent, prejudice, warp.

Carrion Dead or decaying flesh.

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 Subjects are followed backward in time, from effects 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 The use of screening tests to search for disease among a clinician's own patients, who are consulting for unrelated symptoms. Typically, every animal is sampled and the objective is to identify the affected individual.

Case report Detailed presentation of a single case or a handful of cases (<10); may be either cross-sectional or longitudinal.

Case series Cross-sectional study with no defined population and no comparison group. Censored observations. Data on patients with incomplete follow-up.

Climatograph Graphs in which total precipitation is plotted against mean temperature for each month, and the resultant points are joined in a closed curve.

Clinical course of disease The progression of disease once it has come under medical care (compare with natural history of disease).

Clinical epidemiology 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.

Coefficient of determination (r^2) The square of the correlation coefficient. A measure of closeness of fit of the data to the linear regression line. The value for r^2 expresses the amount of variation in the data that are accounted for by the linear relationship between two variables and may take any value between 0 and 1. 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.

Cohort 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 (see 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 (Benenson, 1985).

Communicable period The time or times 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 invertebrate vectors (Benenson, 1985).

Compliance The proportion of individuals (or their owners) that adhere to the 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.

Conditional likelihood of a disease An estimate of likelihood that the observed morphologic findings would occur in a disease. Basically an expression of sensitivity data.

Confidence interval The theoretical range over which there is a specified probability (usually 95%) of including the true value.

Confounding variable (1) A variable that is related to two factors of interest (e.g., disease state and degree of exposure to some agent in a case-control study; treatment assignment and outcome in a clinical trial) that falsely obscures or accentuates the relationship between the factors. (2) A baseline variable in a clinical trial that influences the outcome and that has a different distribution in the treatment groups being compared.

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.

Correlation coefficient (r) The square root of the coefficient of determination. A measure of the degree of linear association between two variables. 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 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.

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 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.

Crude death rate Number of deaths during an outbreak/mean population during the outbreak period.

Cyclical changes Increases or decreases in rates (such as disease incidence) developing at intervals longer than a year.

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 test 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 (White, 1986). Diagnostic testing begins with diseased individuals.

Dissemination See mode of spread.

Double counting A form of multiple testing bias that occurs when interpretation of a test finding is based, in part, on prior test findings. This may occur when two or more tests really measure the same thing (such as the same class of antibody), or when two or more specialists (as clinician and pathologist) interpret findings from the same clinical case.

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 "host, agent and environment triad."

Effectiveness A measure of how well a treatment works among those to whom it is offered (compare with efficacy).

Efficacy The power to produce effects or intended results. 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 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.

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.

Evapotranspiration The combined evaporation from the soil surface and transpiration from plants. It is the reverse of precipitation, since it represents the transport of water from the earth back to the atmosphere.

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).

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.

Extrinsic risk factors Risk factors that are not properties of the host, i.e., agent and environment.

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 or not. 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 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.

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 of the event. An incidence of this type is expressed as the number of new cases per total number of animal days or years at risk.

Interval 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 rate (basic reproductive rate) The number of secondary infections produced by one case in a totally susceptible population.

Intrinsic risk factors Risk factors that are properties of the host.

Irregular variation Reflects random variation in disease occurrence among individuals in a population.

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. The likelihood ratio for a positive 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).

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 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.

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.

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.

Natural history of disease The evolution of disease without medical intervention (compare with clinical course of disease).

Negative correlation See correlation coefficient.

Nominal 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 judgement, owner preference).

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, e.g., the stronger the association between exposure and disease, the higher the odds ratio.

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. For analytic purposes ordinal data may be treated as continuous or discrete variables.

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 at the same time. The net effect of parallel testing is to ask the patient to prove that it is healthy.

Parenteral Not through the alimentary canal, i.e., such as subcutaneous, intramuscular, intradermal, intravenous, etc.

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.

Period prevalence Number of cases (old and new) detected over a time period/number of animals examined over the same time period.

Placebo In clinical trials, an intervention that is indistinguishable from the active treatment, but does not possess its specifically active component.

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.

Posterior likelihood of a disease The product of the prior likelihood and conditional likelihood of a disease. Also known as the revised likelihood of a disease.

Power of a test The probability that a trial will find a statistically significant difference when a difference really exists. A powerful test 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.

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.

Prior likelihood of a disease A numerical estimate of the probability of any disease in a cohort of patients identical to the one in question. It is based in part on the combination of signs and symptoms, and in part on the prevalence of the condition in the population.

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.

Real prevalence The prevalence of disease estimated through use of an appropriate gold standard (compare with apparent prevalence).

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.

Revised likelihood of a disease See posterior likelihood of a disease.

Risk factors Factors that are associated with an increased likelihood of acquiring disease.

Route of infection The route by which an etiologic agent gains access to the body of a susceptible individual.

Screening The presumptive identification of unrecognized disease or defect in apparently healthy populations.

Seasonal fluctuations Regular changes in incidence rates with periods shorter than a year.

Secular trends Overall long-term rises or declines in incidence rate that occur gradually over long periods of time.

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. 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.).

Specific seasonals A ratio in which the observed monthly disease incidence rate is divided by the 12-month moving average value centered on the middle of that month.

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.

Specific rate A rate for a specific subgroup of a population of interest (example: 3-5 year age group). Compare with crude rate.

Sporadic disease A disease which occurs rarely and without regularity in a population unit (compare with endemic and epidemic).

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, and so usually they are mathematically adjusted to have the same proportions as a standard population.

Statistically significant 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 is based on our own observations and those of the owner.

Survival cohort A group of patients who are assembled at various times in the course of their disease, rather than at the beginning, and who are then observed for a period of time to see what happens to them. Generally not considered a true cohort (see cohort).

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.

Synanthropic Together with or accompanying human beings.

Type I error See alpha error.

Type II error See beta error.

Transmissible (communicable) infection An infection that can be passed from infected to susceptible animals.

Typical seasonals Indices of the amount of variation attributable to seasonal influences obtained by averaging (by mean or median) the specific seasonals for each month.

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 (Benenson, 1985).

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.

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).

REFERENCES

Abbitt, B., Murphy, M.J., Ray, A.C., Reagor, J.C., Eugster, A.K., Gayle, L.G., Whitford, H.W., Sutherland, R.J., Fiske, R.A., and Pusok, J. 1984. Catastrophic death losses in a dairy herd attributed to type D botulism. *J.A.V.M.A.* 185:798-801.

Acha, P.N. and Szyfres, B. 1980. *Zoonoses and Communicable Diseases Common to Man and Animals.* Pan American Health Organization. Washington, D.C. 700 pp.

American Heartworm Society. 1993. American Heartworm Society recommended procedures for the diagnosis and management of heartworm (*Dirofilaria immitis*) infection. Am. Heartworm Soc. Bulletin. 19(1):1-8.

Anderson, R.M. and May, R.M. 1982. *Population Biology of Infectious Diseases.* Springer-Verlag, New York. 315 pp.

Anonymous. 1980. Opie on the heart. *Lancet.* 1:692.

Anonymous. 1985a. Tracking swine disease through the packing plant. *Pig American.* January, 1985, pp. 20-25.

Anonymous. 1985b. Study finds hog respiratory disease costing producers $200 million yearly. *DVM Magazine.* October, 1985, p. 57.

Anonymous. 1987. *Swine Pseudorabies Eradication Guidelines,* second edition. Pseudorabies Committee, Livestock Conservation Institute, South St. Paul, MN. 12 pp.

AVMA. 1984. FDA spells out extra-label use guidelines. *J.A.V.M.A.* 185:950.

AVMA. 1986. Student enrollment (1985-1986) favors women. *J.A.V.M.A.* 188:573-575.

Benenson, A.S. 1985. *Control of Communicable Diseases in Man,* 14th edition. American Public Health Association. 485 pp.

Bennett, R.M. 1992. The use of 'economic' quantitative modelling techniques in livestock health and disease-control decision making: a review. *Prev. Vet. Med.* 13:63-76.

Bishopp, F.C. and Trembley, H.L. 1945. Distribution and hosts of certain North American ticks. *J. Parasitol.* 31:1-54.

Blair, A. and Hayes, H.M., Jr. 1982. Mortality patterns among U.S. veterinarians, 1947-1977: an expanded study. *Int. J. Epidemiol.* 11:391-397.

Boon, G.D. and Rebar, A.H. 1984. *Veterinary Values,* second edition. Schering Corp., Kenilworth, NJ, p. 7.

Bowman, G.L., Hueston, W.D., Boner, G.J., Hurley, J.J., and Andreas, J.E. 1986. *Serratia liquefaciens* mastitis in a dairy herd. *J.A.V.M.A.* 189:913-915.

Bronson, R.T. 1982. Variation in age at death of dogs of different sexes and breeds. *Am. J. Vet. Res.* 43:2057-2059.

Bureau of Census. 1978. Expectation of life and mortality rates, by race, age, and sex: 1976. Statistical Abstract of the United States. U.S. Department of Commerce.

Burnett, C.D. 1989. Bat rabies in Illinois: 1965-1986. *J. Wildl. Dis.* 25:10-19.

Carter, J.D., Hird, D.W., Farver, T.B., and Hjerpe, C.A. 1986. Salmonellosis in hospitalized horses: seasonality and case fatality rates. *J.A.V.M.A.* 188:163-167.

CDC. 1986a. Rocky Mountain spotted fever -- United States, 1985. *MMWR.* 35(Apr. 18, 1986; No. 15):247-249.

CDC. 1986b. Group-A, -B hemolytic *Streptococcus* skin infections in a meat-packing plant -- Oregon. *MMWR.* 35(Oct. 10, 1986; No. 40):629-630.

Center, S.A., Baldwin, B.H., Dillingham, S., Erb, H.N., and Tennant, B.C. 1986. Diagnostic value of serum gamma-glutamyl transferase and alkaline phosphatase activities in hepatobiliary disease in the cat. *J.A.V.M.A.* 188:507-510.

Childs, J.E. and Ross, L. 1986. Urban cats: characteristics and estimation of mortality due to motor vehicles. *Am. J. Vet. Res.* 47:1643-1648.

Cochran, W.G. 1977. *Sampling Techniques*, third edition. Wiley & Sons, New York. 428 pp.

Cole, J.R., Hall, R.F., Gosser, H.S., Hendricks, J.B., Pursell, A.R., Senne, D.A., Pearson, J.E., and Gipson, C.A. 1986. Transmissibility and abortogenic effect of equine viral arteritis in mares. *J.A.V.M.A.* 189:769-771.

Collins, M.T. and Sockett, D.C. 1993. Accuracy and economics of the USDA-licensed enzyme-linked immunosorbent assay for bovine paratuberculosis. *J.A.V.M.A.* 203:1456-1463.

Collins, M.T., Sockett, D.C., Goodger, W.J., Conrad, T.A., Thomas, C.B., and Carr, D.J. 1994. Herd prevalence and geographic distribution of, and risk factors for, bovine paratuberculosis in Wisconsin. *J.A.V.M.A.* 204:636-641.

Courtney, C.H., Zeng, Q.Y, and Bean, E.S. 1988. 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.* 24:27-32.

Courtney, C.H. and Cornell, J.A. 1990. Evaluation of heartworm immunodiagnostic tests. *J.A.V.M.A.* 197:724-729.

Courtney, C.H., Zeng, Q.Y., and Tonelli, Q. 1990. Sensitivity and specificity of the CITE heartworm antigen test and a comparison with the DiroChek heartworm antigen test. *J. Am. Anim. Hosp. Assoc.* 26:623-628.

Crawford, T.B. and Adams, D.S. 1981. Caprine-arthritis-encephalitis: clinical features and presence of antibody in selected goat populations. *J.A.V.M.A.* 178:713-719.

Crow, S.E. 1985. Usefulness of prognoses: qualitative terms vs. quantitative designations. *J.A.V.M.A.* 187:700-703.

DiBartola, S.P., Rutgers, H.C., Zack, P.M., and Tarr, M.J. 1987. Clinicopathologic findings associated with chronic renal disease in cats: 74 cases (1973-1984). *J.A.V.M.A.* 190:1196-1202.

Dohoo, I.R. and Waltner-Toews, D. 1985a. Interpreting clinical research. Part I. General considerations. *Compend. Cont. Educ. Pract. Vet.* 7:S473-S478.

Dohoo, I.R. and Waltner-Toews, D. 1985b. Interpreting clinical research. Part II. Descriptive and experimental studies. *Compend. Contin. Educ. Pract. Vet.* 7:S513-S520.

Dohoo, I.R. and Waltner-Toews, D. 1985c. Interpreting clinical research. Part III. Observational studies and interpretation of results. *Compend. Contin. Educ. Pract. Vet.* 7:S605-S613.

Dohoo, I.R., Morris, R.S., Martin, S.W., Perry, B.D., Bernardo, T., Erb, H., Thrusfield, M., Smith, R., and Welte, V.R. 1994. Epidemiology (letter). *Nature.* 368:284.

Dorland's Illustrated Medical Dictionary. 1981. Twenty-sixth edition. W.B. Saunders, Philadelphia. 1485 pp.

Dubensky, R.A. and White, M.E. 1983. The sensitivity, specificity and predictive value of total plasma protein in the diagnosis of traumatic reticuloperitonitis. *Can. J. Comp. Med.* 47:241-244.

Dubey, J.P., Murrell, K.D., Hanbury, R.D., Anderson, W.R., Doby, P.B., and Miller, H.O. 1986. Epidemiologic findings on a swine farm with enzootic toxoplasmosis. *J.A.V.M.A.* 189:55-56.

East, N.E., Rowe, J.D., Madewell, B.R., and Floyd, K. 1987. Serologic prevalence of caprine arthritis-encephalitis virus in California goat dairies. *J.A.V.M.A.* 190:182-186.

Evans, G.O. 1987. Plasma lactate measurements in healthy Beagle dogs. *Am. J. Vet. Res.* 48:131-132.

Evermann, J.F., DiGiacomo, R.F., Ferrer, J.F., and Parish, S.M. 1986. Transmission of bovine leukosis virus by blood inoculation. *Am. J. Vet. Res.* 47:1885-1887.

Fagan, T.J. 1975. Nomogram for Bayes's theorem. *N. Engl. J. Med.* (letter) 293:257.

Fertig, D.L. and Dorn, C.R. 1985. *Taenia saginata* cysticercosis in an Ohio cattle feeding operation. *J.A.V.M.A.* 186:1281-1285.

Fetrow, J., Madison, J.B., and Galligan, D. 1985. Economic decisions in veterinary practice: a method for field use. *J.A.V.M.A.* 186:792-797.

Fettman, M.J. 1987. Evaluation of the usefulness of routine microscopy in canine urinalysis. *J.A.V.M.A.* 190:892-896.

Fletcher, R.H., Fletcher, S.W., and Wagner, E.H. 1982. *Clinical Epidemiology – The Essentials.* Williams & Wilkins, Baltimore, MD. 223 pp.

Friedmann, C.T.H., Spiegel, E.R., Aaron, E., and McIntyre, R. 1971 *CDC reports.* p10.

Geller, S.A. 1983. Autopsy. *Scientific American.* 248:124-136.

Gingerich, D.A., Rourke, J.E., Chatfield, R.C., and Strom, P.W. 1985. Butorphanol tartrate: a new analgesic to relieve the pain of equine colic. *Vet. Med.* 80(8):72-77.

Goodger, W.J. and Skirrow, S.Z. 1986. Epidemiologic and economic analyses of an unusually long epizootic of trichomoniasis in a large California dairy herd. *J.A.V.M.A.* 189:772-776.

Gordis, L. 1980. Challenges to epidemiology in the coming decade. *Am. J. Epidemiol.* 112:315-321.

Halpin, B. 1975. *Patterns of Animal Disease.* Williams & Wilkins, Baltimore, MD. 184 pp.

Hanbury, R.D., Doby, P.B., Miller, H.O., and Murrell, K.D. 1986. Trichinosis in a herd of swine: cannibalism as a major mode of transmission. *J.A.V.M.A.* 188:1155-1159.

Hannah, H.W. 1985. Promising a result. *J.A.V.M.A.* 186:1166.

Hardy, W.D., Jr., McClelland, A.J., Zuckerman, E.E., Hess, P.W., Essex, M., Cotter, S.M., MacEwen, E.G., and Hayes, A.A. 1976. Prevention of the contagious spread of feline leukaemia virus and the development of leukaemia in pet cats. *Nature.* 263:326-328.

Hart, B.L. and Miller, M.F. 1985. Behavioral profiles of dog breeds. *J.A.V.M.A.* 186:1175-1180.

Hawkins, E.C. and Murphy, C.J. 1986. Inconsistencies in the absorptive capacities of the Schirmer tear test strips. *J.A.V.M.A.* 188:511-513.

Herd, R.P. and Heider, L.E. 1985. Control of nematodes in dairy heifers by prophylactic treatments with albendazole in the spring. *J.A.V.M.A.* 186:1071-1074.

Hicks, C.R., Eberhart, R.J., and Sischo, W.M. 1994. Comparison of microbiologic culture, an enzyme-linked immunosorbent assay, and determination of somatic cell count for diagnosing *Staphylococcus aureus* mastitis in dairy cows. *J.A.V.M.A.* 204:255-260.

Hildebrandt, P.K., Huxsoll, D.L., Walker, J.S., Nims, R.M., Taylor, R., and Andrews, M. 1973. Pathology of canine ehrlichiosis (tropical canine pancytopenia). *Am. J. Vet. Res.* 34:1309-1320.

Hird, D.W., Casebolt, D.B., Carter, J.D., Pappaioanou, M., and Hjerpe, C.A. 1986. Risk factors for salmonellosis in hospitalized horses. *J.A.V.M.A.* 188:173-177.

Hoblet, K.H., Miller, G.Y., and Bartter, N.G. 1987. Economic assessment of a pseudorabies epizootic, breeding herd removal/repopulation, and downtime in a commercial swine herd. *J.A.V.M.A.* 190:405-409.

Holden, C. 1985. IOM sees need for autopsy policy. *Science.* 229:539.

Hoskins, J.D., Hribernik, T.N., and Kearney, M.T. 1985. Complications following thiacetarsamide sodium therapy in Louisiana dogs with naturally-occurring heartworm disease. *Cornell Vet.* 75:531-539.

House, J.A. and Baker, J.A. 1968. Comments on combination vaccines for bovine respiratory diseases. *J.A.V.M.A.* 152:893-894.

Houston, D.M. 1984. Impact of technology and funding on meat and poultry inspection. *J.A.V.M.A.* 185:1505-1507.

Huntsberger, D.V. and Billingsley, P. 1973. *Elements of Statistical Inference*, third edition. Allyn & Bacon, Boston, MA. 349 pp.

Jeglum, K.A., de Guzman, E., and Young, K.M. 1985. Chemotherapy of advanced mammary adenocarcinoma in 14 cats. *J.A.V.M.A.* 187:157-160.

Kasari, T.R. and Naylor, J.M. 1985. Clinical evaluation of sodium bicarbonate, sodium L-lactate, and sodium acetate for the treatment of acidosis in diarrheic calves. *J.A.V.M.A.* 187:392-397.

Kassirer, J.P., Moskowitz, A.J., Lau, J., and Pauker, S.G. 1987. Decision analysis: a progress report. *Annals of Internal Medicine.* 106:275-291.

King, L.J. 1985. Unique characteristics of the National Animal Disease Surveillance System. *J.A.V.M.A.* 186:35-39.

Kleinbaum, D.G. and Kleinbaum, A. 1976. *Adjusted rates. The direct rate. A self-instructional program.* Publication 122-00-004. Health Sciences Consortium, Inc., Chapel Hill, NC. 54 pp.

Koterba, A., Torchia, J., Silverthorne, C., Ramphal, R., Merritt, A.M., and Manucy, J. 1986. Nosocomial infections and bacterial antibiotic resistance in a university equine hospital. *J.A.V.M.A.* 189:185-191.

Kunkle, G.A. and Milcarsky, J. 1985. Double-blind flea hyposensitization trial in cats. *J.A.V.M.A.* 186:677-680.

Lebeau, A. 1953. L'age du chien et celui de l'homme. Essai de statistique sur la mortalite canine. *Bull. Acad. Vet. France.* 26:229-232.

Levine, N.D. 1963. Weather, climate, and the bionomics of ruminant nematode larvae. *Adv. Vet. Sci.* 8:215-261.

Levine, N.D. 1965. Bioclimatographs, evapotranspiration, soil moisture data and the free-living stages of ruminant nematodes and other disease agents. *Theoretical Questions of Natural Foci of Diseases,* B. Rosicky and K. Heyberger (eds), pp. 455-461. Czechoslovak Academy of Sciences.

Lewis, G.E., Jr., Ristic, M., Smith, R.D., Lincoln, T., and Stephenson, E.H. 1977. The brown dog tick *Rhipicephalus sanguineus* and the dog as experimental hosts of *Ehrlichia canis*. *Am. J. Vet. Res.* 38:1953-1955.

Little, C. and Hilbert, B. 1987. Pelvic fractures in horses: 19 cases (1974-1984). *J.A.V.M.A.* 190:1203-1206.

Losonsky, J.M. and Kneller, S.K. 1988. Variable locations of nutrient foramina of the proximal phalanx in forelimbs of Standardbreds. *J.A.V.M.A.* 193:671-673.

MacMahon, B. and Pugh, T.F. 1970. *Epidemiology: Principles and Methods.* Little, Brown & Co., Boston. 376 pp.

Madison, J.B., Fetrow, J., and Galligan D. 1984. Economic decisions in food animal practice: To treat or not to treat? *J.A.V.M.A.* 185:520-521.

Martin, S.W., Meek, A.H., and Welleberg, P. 1987. *Veterinary Epidemiology. Principles and Methods.* Iowa State University Press, Ames, IA. 343 pp.

Matus, R.E., Leifer, C.E., MacEwen, E.G., and Hurvitz, A.I. 1986. Prognostic factors for multiple myeloma in the dog. *J.A.V.M.A.* 188:1288-1292.

May, R.M. 1983. Parasitic infections as regulators of animal populations. *Am. Scientist.* 71:36-45.

McClelland, A.J., Hardy, W.D., Jr., and Zuckerman, E.E. 1980. Prognosis of healthy feline leukemia virus infected cats. In, W.D. Hardy, Jr., M. Essex, and A.J. McClelland (eds), *Feline Leukemia Virus.* Elsevier, New York, pp. 121-126.

McCoy (APHIS) 1985. Personal communication.

McNeil, B.J., Keeler, E., and Adelstein, S.J. 1975. Primer on certain elements of medical decision making. *N. Engl. J. Med.* 293:211-226.

Merkal, R.S., Whipple, D.L., Sacks, J.M., and Snyder, G.R. 1987. Prevalence of *Mycobacterium paratuberculosis* in ileocecal lymph nodes of cattle culled in the United States. *J.A.V.M.A.* 190:676-680.

Mills, K.W. and Kelly, B.L. 1986. Antibiotic susceptibilities of swine to *Salmonella* isolants from 1979 to 1983. *Am. J. Vet. Res.* 47:2349-2350.

Moore, B.R., Reed, S.M., Biller, D.S., Kohn, C.W., and Weisbrode, S.E. 1994. Assessment of vertebral canal diameter and bony malformations of the cervical part of the spine in horses with cervical stenotic myelopathy. *Am. J. Vet. Res.* 55:5-13.

Morrison, R.B. and Joo, H.S. 1985. Prenatal and preweaning deaths caused by pseudorabies virus and porcine parvovirus in a swine herd. *J.A.V.M.A.* 187:481-483.

Morton, R.F. and Hebel, J.R. 1979. *A Study Guide to Epidemiology and Biostatistics.* University Park Press, Baltimore, MD. 153 pp.

Murrell, K.D., Anderson, W.R., Schad, G.A., Hanbury, R.D., Kazacos, K.R., Gamble, H.R., and Brown, J. 1986. Field evaluation of the enzyme-linked immunosorbent assay for swine trichinosis: efficacy of the excretory-secretory antigen. *Am. J. Vet. Res.* 47:1046-1049.

Owen, L.N. (ed.) 1980. *TNM Classification of Tumours in Domestic Animals*, first edition. WHO, Geneva.

Oxender, W.D., Newman, L.E., and Morrow, D.A. 1973. Factors influencing dairy calf mortality in Michigan. *J.A.V.M.A.* 162:458-460.

Padgett, G. 1985. New research unlocking genetic disease mysteries. *DVM.* 16:30-32.

Palmer, J.E., Whitlock, R.H., and Benson, C.E. 1986. Equine ehrlichial colitis (Potomac horse fever): recognition of the disease in Pennsylvania, New Jersey, New York, Ohio, Idaho, and Connecticut. *J.A.V.M.A.* 189:197-199.

Pauker, S.G. and Kassirer, J.P. 1980. The threshold approach to clinical decision making. *N. Engl. J. Med.* 302:1109-1117.

Pauker, S.G. and Kassirer, J.P. 1987. Decision analysis. *New Engl. J. Med.* 316:250-258.

Payne, J.M., Dew, S.M., Manston, R., and Faulks, M. 1970. The use of a metabolic profile test in dairy herds. *Vet. Rec.* 87:150-157.

Pion, P.D., Kittleson, M.D., Thomas, W.P., Delellis, L.A., and Rogers, Q.R. 1992. Response of cats with dilated cardiomyopathy to taurine supplementation. *J.A.V.M.A.* 201:275-284.

Ransohoff, D.F. and Feinstein, A.R. 1978. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N. Engl. J. Med.* 299:926-930.

Reif, J.S., Maguire, T.G., Kenney, R.M., and Brodey, R.S. 1979. A cohort study of canine testicular neoplasia. *J.A.V.M.A.* 175:719-723.

Risco, C.A., Reynolds, J.P., and Hird, D. 1984. Uterine prolapse and hypocalcemia in dairy cows. *J.A.V.M.A.* 185:1517-1519.

Ruble, R.P. and Hird, D.W. 1993. Congenital-Abnormalities in Immature Dogs from a Pet Store – 253 Cases (1987-1988). *J.A.V.M.A.* 202:633-636.

Sackett, D.L. 1992. A primer on the precision and accuracy of the clinical examination. *J.A.M.A.* 267:2638-2644.

Sackett, D.L., Haynes, R.B., Guyatt, G.H., and Tugwell, P. 1991. *Clinical Epidemiology. A Basic Science for Clinical Medicine*, second edition. Little, Brown & Co., Boston. 441 pp.

Sandlow, L.J., Hammett, W.H., and Bashook, P.G. 1974. *Problem Oriented Medical Records. Guidelines for Format and Forms.* Michael Reese Medical Center, Chicago, IL. 41 pp.

Sanford, S.E. 1987. Enteric cryptosporidial infection in pigs: 184 cases (1981-1985). *J.A.V.M.A.* 190:695-698.

Scavelli, T.D., Patnaik, A.K., Mehlhaff, C.J., and Hayes, A.A. 1985. Hemangiosarcoma in the cat: retrospective evaluation of 31 surgical cases. *J.A.V.M.A.* 187:817-819.

Schick, R.O. and Fadok, V.A. 1986. Responses of atopic dogs to regional allergens: 268 cases (1981-1984). *J.A.V.M.A.* 189:1493-1496.

Schnurrenberger, P.R., Martin, R.J., and Walker, J.F. 1972. Characteristics of veterinarians in Illinois. *J.A.V.M.A.* 160:1512-1521.

Schnurrenberger, P.R., Walker, J.F., and Martin, R.J. 1975. *Brucella* infections in Illinois veterinarians. *J.A.V.M.A.* 167:1084-1088.

Schwabe, C.W. 1984. Analytical epidemiology and veterinary economics. *Veterinary Medicine and Human Health*, third edition. Williams & Wilkins, Baltimore, MD, pp. 430-447.

Schwabe, C.W., Riemann, H.P., and Franti, C.E. 1977. *Epidemiology in Veterinary Practice.* Lea & Febiger, Philadelphia. 303 pp.

Schwartz, W.B., Wolfe, H.J., and Pauker, S.G. 1981. Pathology and probabilities. A new approach to interpreting and reporting biopsies. *N. Engl. J. Med.* 305:917-923.

Schwartz, B.S., Goldstein, M.D., Ribeiro, J.M.C., Schulze, T.L., and Shahied, S.I. 1989. Antibody testing in Lyme disease. A comparison of results in four laboratories. *J.A.V.M.A.* 262:3431-3434.

Sharp, V.F. 1979. *Statistics for the Social Sciences.* Little, Brown & Co., Boston, MA. 381 pp.

Shott, S. 1985. Statistics in veterinary research. *J.A.V.M.A.* 187:138-141.

Smith, M.M., Vasseur, P.B., and Morgan, J.P. 1985. Clinical evaluation of dogs after surgical and nonsurgical management of osteochondritis dissecans of the talus. *J.A.V.M.A.* 187:31-35.

Smith, R.D. 1977. *Ehrlichiae. Parasitic Protozoa*, Vol. IV. J.P. Kreier (ed). Academic Press, NY, pp. 295-328.

Smith, R.D. 1988. Veterinary clinical research: a survey of study designs and clinical issues appearing in a practice journal. *Journal of Veterinary Medical Education* 15(1):2-7.

Smith, R.D. 1993. Decision analysis in the evaluation of diagnostic tests. *J.A.V.M.A.* 203:1184-1192.

Spangler, C., Bech-Nielsen, S., Heider, L.E., and Dorn, C.R. 1992. Interpretation of an enzyme-linked immunosorbent test using different cut-offs between positive and negative samples for diagnosis of paratuberculosis. *Prev. Vet. Med.* 13:197-204.

Spain, J.D. 1982. *BASIC Microcomputer Models in Biology.* Addison-Wesley, p. 114.

Stein, T.E. and Duffy, S.J. 1988. Parity-specific production values for 68 North American swine breeding herds. Proceedings of the 5th Int'l. Symposium on Veterinary Epidemiology and Economics, July 25-29, 1988, Copenhagen, Denmark. *Acta Vet. Scand.* Supplementum 84:522.

Stevens, J.B., Anderson, J.F., Olson, W.G., and Schlotthauer, J.C. 1980. Metabolic profile testing. In, H.E. Amstutz (ed), *Bovine Medicine and Surgery, Vol. 1.* American Veterinary Publications, Santa Barbara, CA, pp. 597-614.

Straus, J.H. 1982. Anemia. In, W.R. Fenner (ed), *Quick Reference to Veterinary Medicine.* J.B. Lippincott Co., Philadelphia, pp. 383-398.

Straw, B. 1985. The practical value of slaughter checks. *Pig American.* January, pp. 20-25.

Straw, B.E., Henry, S.C., and Fleming, S.A. 1985. Interactions of management and animal performance in a swine feedlot. *J.A.V.M.A.* 186:986-988.

Straw, B.E., Backstrom, L., and Leman, A.D. 1986. Examination of swine at slaughter. Part I. The mechanics of slaughter examination and epidemiologic considerations. *Compend. Cont. Educ. Pract. Vet.* 8:S41-S47.

Thomas, D.G., Breslow, N., and Gart, J.J. 1977. Trend and homogeneity analyses of proportions and life table data. *Comp. Biomed. Res.* 10:373-381.

Troutman, C.M. 1988. Veterinary services market for companion animals: summary report. *J.A.V.M.A.* 193:920-922.

Turner, T.A. 1986. Shoeing principles for the management of navicular disease in horses. *J.A.V.M.A.* 189:298-301.

Turrel, J.M. 1987. Intraoperative radiotherapy of carcinoma of the prostate gland in ten dogs. *J.A.V.M.A.* 190:48-52.

USDA, 1981. A guide for accredited veterinarians. APHIS 91-18. U.S. Government Printing Office.

Vasseur, P.B., Paul, H.A., Enos, L.R., and Hirsh, D.C. 1985. Infection rates in clean surgical procedures: a comparison of ampicillin prophylaxis vs a placebo. *J.A.V.M.A.* 187:825-827.

Warner, H.R., Haug, P., Bouhaddou, O., Lincoln, M., Warner, H., Jr., Sorenson, D., Williamson, J.W., and Fan, C. 1988. ILIAD as an expert consultant to teach differential diagnosis. In, R.A. Greenes (ed), Proceedings of the Twelfth Annual Symposium on Computer Applications in Medical Care, November 6-9, 1988, Washington, D.C., pp. 371-376.

Webber, J.J. and Selby, L.A. 1981. Risk factors related to the prevalence of infectious bovine keratoconjunctivitis. *J.A.V.M.A.* 179:823-826.

Whetstone, C.A., Wheeler, J.G., and Reed, D.E. 1986. Investigation of possible vaccine-induced epizootics of infectious bovine rhinotracheitis, using restriction endonuclease analysis of viral DNA. *Am. J. Vet. Res.* 47:1789-1795.

White, M.E. 1986. Evaluating diagnostic test results (letter). *J.A.V.M.A.* 188:1141.

Wise, J.K. 1993. 1991 professional incomes of US veterinarians. *J.A.V.M.A.* 202:210-212.

Zimmerman, W.J. and Zinter, D.E. 1971. The prevalence of trichinosis in swine in the United States. *HSMHA Health Rep.* 86:937-941.

Zweig, M.H. and Campbell, G. 1993. Receiver-Operating Characteristic (ROC) Plots – A Fundamental Evaluation Tool in Clinical Medicine. *Clin. Chem.* 39:561-577.

INDEX

D

E

F

G

H

I

J

K

L

M

N

O

P

R

S

T

U

V

W

Z

2445
ISBN 0-8493-2445-9