



Veterinary Epidemiology

- An Introduction

Dirk U. Pfeiffer

Professor of Veterinary Epidemiology

Epidemiology Division
Department of Veterinary Clinical Sciences,
The Royal Veterinary College,
University of London

September 2002

Postal: Royal Veterinary College, Hawkshead Lane, North Mymms, Hertfordshire AL9 7TA, United Kingdom;
E-mail: pfeiffer@rvc.ac.uk; **Fax:** +44 (1707) 666 346; **Tel:** +44 (1707) 666 342

TABLE OF CONTENTS

BASIC CONCEPTS OF VETERINARY EPIDEMIOLOGY	3
Learning Objectives	3
Challenges to Today's Veterinary Medicine	3
Veterinary Epidemiology	3
Evidence-based Veterinary Medicine (EBVM)	4
Evidence-based Veterinary Medicine in Practice	4
Basic Epidemiological Concepts	5
Causation	9
DESCRIPTIVE EPIDEMIOLOGY	12
Learning Objectives	12
Measurement of Disease Frequency and Production	12
Survival	15
Standardisation of Risk	16
ANALYTICAL EPIDEMIOLOGY	18
Learning Objectives	18
Introduction	18
Epidemiological Studies	19
Concept of Risk	22
Identification of Risk Factors	22
From Association to Inference in Epidemiological Studies	25
SAMPLING OF ANIMAL POPULATIONS	28
Learning Objectives	28
Introduction	28
Sample Size Considerations	34
INTERPRETATION OF DIAGNOSTIC TESTS	37
Learning Objectives	37
Uncertainty and the Diagnostic Process	37
Diagnostic Tests	38
Evaluation and Comparison of Diagnostic Tests	38
Test Performance and Interpretation at the Individual Level	39
Methods for choosing Normal/Abnormal Criteria	43
Likelihood Ratio	44
Combining Tests	46
Decision Analysis	50
EPIDEMIOLOGICAL ANIMAL DISEASE INFORMATION MANAGEMENT	52
Learning Objectives	52
Outbreak Investigation	52
Assessment of Productivity and Health Status of Livestock Populations	57
Theoretical Epidemiology	58
Information Technology and Veterinary Applications	58
RECOMMENDED READING	60
INDEX	61

Basic Concepts of Veterinary Epidemiology

Learning Objectives

At the completion of this topic, you will be able to:

- understand the concepts of the interrelationships between Agent-Host-Environment, interaction of disease determinants, herd immunity, and web of causation.
- use the terms utilised in infectious disease epidemiology such as infection, incubation period, reservoir, vector, pathogenicity, and virulence.

Challenges to Today's Veterinary Medicine

Today, the veterinary profession is confronted with a different set of problems compared with the middle of last century. Now, veterinarians often have to deal with herds or regions remaining diseased after lengthy disease control campaigns. In addition, it is considered necessary to take into account the economic aspects of disease control through the use of benefit/cost analyses of disease control campaigns. Costly multi-factorial *disease complexes* such as mastitis or cattle lameness have become quite common. In addition, old, emerging or new *diseases* with complex aetiologies pose a difficult challenge for the profession. Veterinarians have to respond to the challenges posed by problems such as BSE, FMD and antimicrobial resistance. All these problems require identification, quantification and intensive examination of multiple, directly or indirectly causal, and often interacting, disease determinants. Veterinary epidemiology and evidence-based veterinary medicine provide overlapping sets of tools which can be used to approach these new challenges.

In clinical practice, there is now a daily need for valid, up-to-date information about diagnosis, therapy and prevention. This situation is complicated by the inadequacy of traditional sources for this type of information, since they may be out-of-date (text books), they are frequently wrong (experts!?), it may be ineffective (didactic CPD), too overwhelming in volume, or too

variable in validity. In addition, the clinician has to deal with the disparity between diagnostic skills and clinical judgement (increasing with experience) on the one hand and up-to-date knowledge and clinical performance (declining with time) on the other. The fact that he/she has insufficient time to examine the animal patient and to practice CPD further complicates the situation.

Veterinary Epidemiology

Veterinary epidemiology deals with the investigation of diseases, productivity and animal welfare in populations. It is used to describe the frequency of disease occurrence and how disease, productivity and welfare are affected by the interaction of different factors or determinants. This information is then used to manipulate such determinants in order to reduce the frequency of disease occurrence. Veterinary epidemiology is a holistic approach aimed at co-ordinating the use of different scientific disciplines and techniques during an investigation of disease or impaired productivity or welfare. The field of veterinary epidemiology can be divided into different components as presented in Figure 1. One of its essential foundations is the collection of data, which then has to be analysed using qualitative or quantitative approaches in order to formulate causal hypotheses. As part of the quantitative approach to epidemiological analysis, epidemiological investigations involving field studies or surveys are being conducted and models of epidemiological problems can be developed. The ultimate goal is to control a disease problem, reduce productivity losses and improve animal welfare.

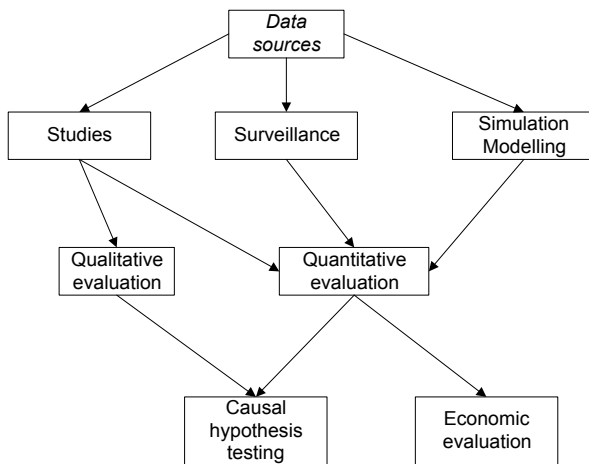


Figure 1: Components of veterinary epidemiology

Evidence-based Veterinary Medicine (EBVM)

In the context of individual animal treatment, the veterinary practitioner has to integrate best research evidence with clinical expertise and patient or owner values. The best research evidence has to be derived from varied sources such as clinically relevant research which may be generated by the basic sciences of veterinary medicine or patient-centred clinical research. The latter will be of often underestimated particular relevance, in that it is research evidence for example describing the accuracy and precision of diagnostic tests, the power of prognostic markers or efficacy and safety of therapeutic, rehabilitative and preventive regimens. The veterinary practitioner is constantly exposed to emerging new evidence which may invalidate previously accepted diagnostic tests and treatments and replace them with new more powerful, more accurate, more efficacious and safer ones. Veterinary epidemiology provides the tools which are used to generate research evidence.

Clinical expertise is the ability to combine clinical skills and past experience to rapidly identify a patient's unique health state, individual risks, benefits of potential interventions, animal welfare needs as well as personal values and expectations of the owner. This is the main focus of the undergraduate training. Best research evidence and clinical expertise have to be applied in the context of the unique

characteristics of the animal patient, the unique preferences, concerns and expectations of the animal owner. The veterinary clinician has to combine all this information in order to achieve the best and economically sustainable clinical outcome and to achieve optimum animal welfare (see Figure 2).

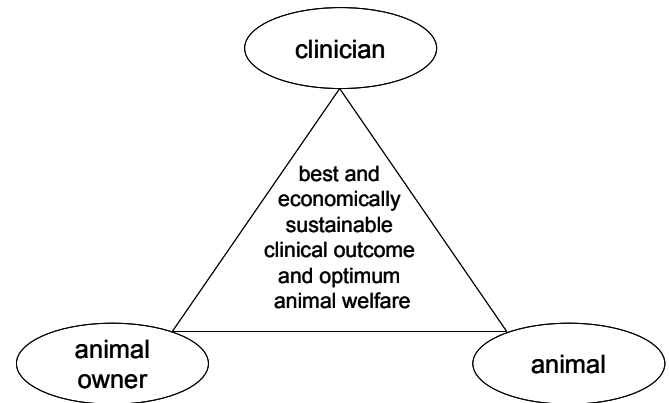


Figure 2: Relationships influencing EBVM

Evidence-based Veterinary Medicine in Practice

An approach incorporating EBVM into clinical practice consists of the following steps:

1. convert need for information into answerable questions
2. track down best evidence
3. critically appraise evidence for validity, impact and applicability
4. integrate critical appraisal with clinical expertise and patient's unique biology, values and circumstances
5. evaluate one's effectiveness and efficiency in steps 1-4 and seek ways for improvement

An EBVM practicing clinician will implement these steps depending on the frequency of a condition. In any case, they will integrate evidence with patient factors (step 4). But there differences with respect to the other steps. With conditions encountered every day, one should work in an appraising mode involving constant searching (step 2) and critical appraisal (step 3) of evidence. In the case of conditions encountered less frequently, the EBVM practitioner will function in a searching mode which represents consultation of critical appraisals done by

others (step 2). And with conditions encountered very infrequently, they work in a replicating mode where they blindly seek, accept and apply recommendations received from authorities in specific branch of veterinary medicine. Unfortunately, this is also the clinical practice mode which is typically applied by students, recent graduates and residents. It also means that one does not consider whether the source of advice operates in an appraising or opinion-based mode, and it is therefore difficult to assess whether the advice is effective, useless or even harmful.

Basic Epidemiological Concepts

The basis for most epidemiological investigations is the assumption that disease does not occur in a random fashion, because one of their main objectives is to identify causal relationships between potential risk factors and outcomes such as disease or productivity losses. Both types of losses are assumed to be influenced by multiple, potentially interacting factors.

Epidemiological investigations focus on general population and disease aspects as well as on causation. In this context, the spatial as well as the temporal dimension of disease occurrence is important. Population parameters which have to be investigated include the health status of the population and factors that are related to health status such as fertility, fecundity, immigration and emigration.

These parameters not only affect the population numerically, but also affect its herd immunity and basic characteristics, such as age structure. Disease within a population is investigated with respect to the possible states of health individuals could be in, such as death, clinical or subclinical disease or health. In individuals, disease is defined as a state of bodily function or form that fails to meet the expectations of the animal owner or society. In populations, it manifests itself through productivity deficits or lack of quality survivorship. Quantitative differences in the manifestation of infectious disease within populations have been described using the

analogy of an *iceberg* (see Figure 3). It assumes that typically a substantial number of animals which were exposed to infection remain uninfected and these represent the base of the *iceberg*. These animals could be susceptible to infection in the future or develop immunity as a consequence of past exposure. Another group of animals may become infected, but has not developed clinical disease. This group of animals may always remain in this category, or could at some stage develop clinical disease depending on the influence of different factors including for example environmental stress. The tip of the *iceberg* includes animals with different manifestations of clinical disease. The ability of animals within these different groups to transmit infection becomes a very important factor in the epidemiology of an infectious disease.

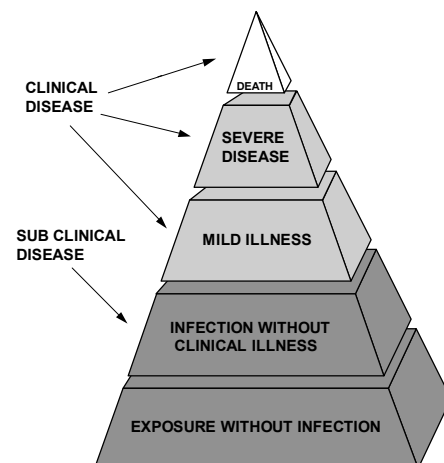


Figure 3: The iceberg concept of disease

Temporal patterns of disease can be broadly categorised into *epidemic* and *endemic* disease. Epidemics are defined as disease occurrence which is higher than expected, whereas endemic disease describes the usual frequency of disease or constant presence of disease. Pandemic disease occurrence refers to widespread epidemics affecting a large proportion of the population and possibly many countries. Sporadic disease occurrence is characterised by situations with single cases or clusters of cases of disease which are normally not present in an area. Temporal patterns are presented graphically using epidemic curves. These are bar charts showing the number of new cases on the

vertical axis and time on horizontal axis (see Figure 4). The shape of the curve can be used to develop hypotheses as to the potential cause of the disease and its epidemiological characteristics.

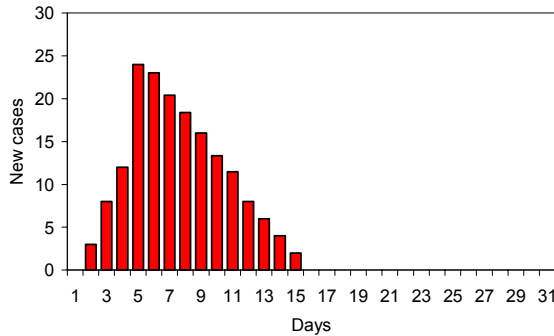


Figure 4: Epidemic curve

Clustering of disease occurrence in time can be described as *short-term variation* as in the case of classical epidemics, *periodic or seasonal variation* such as in the case of human leptospirosis in the U.S.A. (see Figure 5) and *long-term variation* such as with reported wildlife and dog rabies in the U.S.A. (Figure 6).

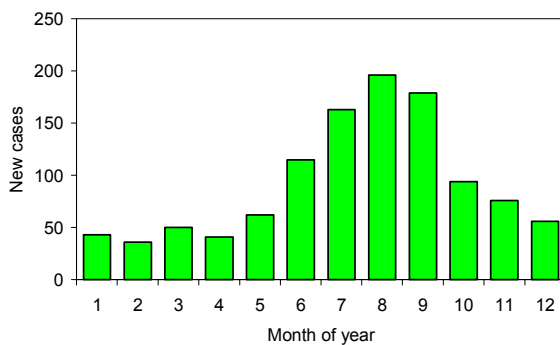


Figure 5: Seasonal occurrence of leptospirosis in humans in USA

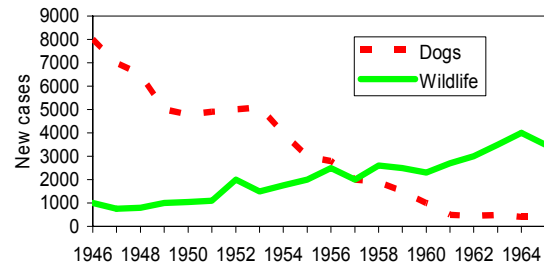


Figure 6: Temporal occurrence of rabies in wildlife and dogs in USA

Figure 7 shows examples of the four standard types of curves of disease occurrence. In the case of a *propagating epidemic*, disease could have been introduced through a single source, and subsequently have been transmitted from infected animals to other susceptible ones within the same population. With *sporadic* disease occurrence only a small number of cases are observed during a short period of time which would infer that the disease process is not infectious under the prevailing conditions. In the case of a *point epidemic* a large number of cases have been observed during a relatively short period of time, but the disease disappears after that time. *Endemic* disease occurrence refers to the appearance of cases at all times.

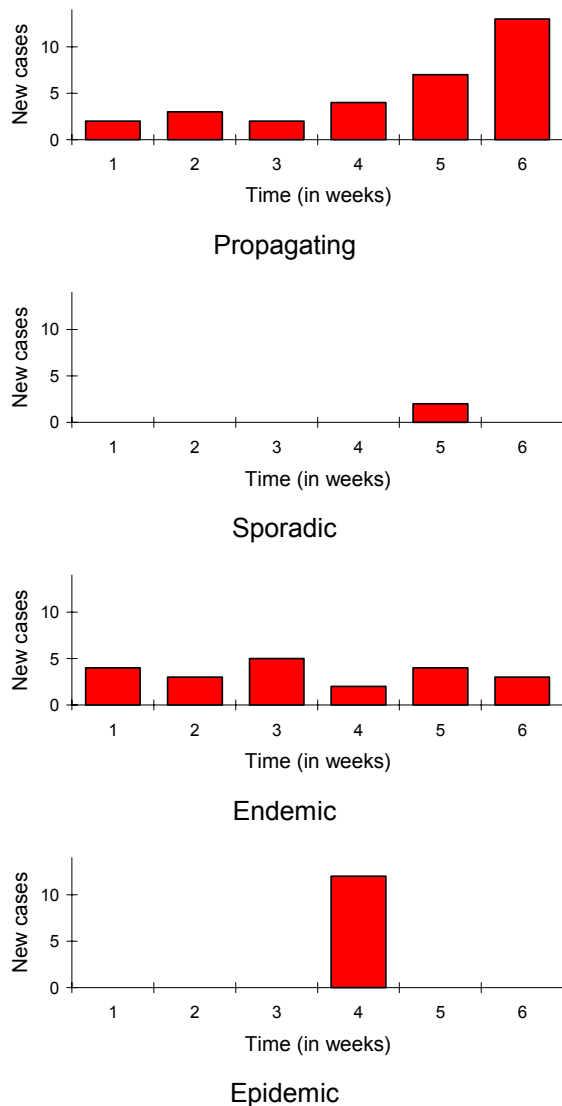


Figure 7: Standard types of temporal patterns of disease occurrence

Disease occurrence can also be characterised through its *spatial pattern* which is typically the consequence of environmental factors differing between locations. Spatial patterns can result from variation between regions and countries, variation within countries or simply local patterns. With the advent of computerised mapping software these types of analyses have become much more accessible to epidemiologists. Figure 8 shows the locations used by wild possums (*Trichosurus vulpecula* Kerr) infected with *Mycobacterium bovis* draped over a three-dimensional representation of the study area. It indicates that locations used by clinically diseased possums are clustered in space.

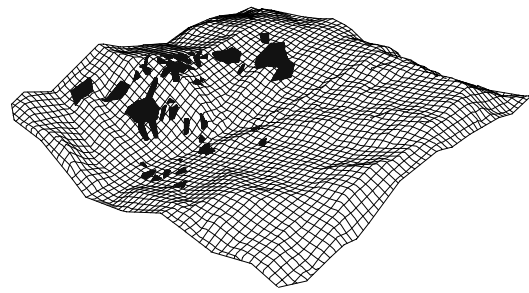


Figure 8: Occurrence of tuberculosis in wild possums during a longitudinal field study (draped over a digital terrain model of the study site)

The concept of *causation* links potential risk factors with the occurrence of disease or impaired productivity. Knowledge of these factors potentially allows control or eradication of disease or increase in productivity. Investigations into cause-effect relationships become useful in populations where not every individual is affected by the problem under study. The objective is then to measure factors describing variation within husbandry systems subject to economic, social, physical and biologic parameters. These factors are also called determinants of health and disease. They can be factors from one or more of these parameter groups, and as risk factors, they may alter the nature or frequency of disease or impaired productivity. *Determinants of disease* include any factor or variable, which can affect the frequency of disease occurrence in a population. These can be of an *intrinsic* nature such as physical or physiological characteristics of the host or disease agent, or *extrinsic* such as environmental influences or interventions by man. *Intrinsic factors* include disease agents which can be living (viruses, bacteria etc.) or non-living (heat and cold, water etc.). *Intrinsic determinants* of living disease agents include *infection* which refers to the invasion of a living organism by another living organism, *infectivity* which is the ability of an agent to establish itself in a host (ID_{50} = numbers of agents required to infect 50% of exposed susceptible animals under controlled conditions) and *pathogenicity* (or virulence) which is the ability of an agent to produce disease in a range of hosts under a range of

environmental conditions. *Virulence* is a measure of the severity of disease caused by a specific agent and is commonly quantified using the LD₅₀ (= numbers of agents required to kill 50% of exposed susceptible animals under controlled conditions). A major component of the epidemiology of any infectious process is the relationship between host and agent. It is characterised as dynamic but there will often be a balance between resistance mechanisms of the host, infectivity and virulence of the agent. The agent increases its survival by increasing its infectivity and decreasing its pathogenicity as well as through shorter generation intervals. A *carrier state* is characterised by an infected host who is capable of dissemination of the agent, but typically does not show evidence of clinical disease. This condition is also called a *true carrier state*. *Incubatory carriers*, on the other hand, are infected, disseminate, but are in the pre-clinical stage. *Convalescent carriers* are infected, disseminate and are in the post-clinical stage. The term *antigenic variation* refers to biological situations where an agent evades the host defence by changing its antigenic characteristics. An example is trypanosomiasis where during a single infection multiple parasitaemias are occurring with antigenetically different trypanosomes in each of them. The *incubation period* is defined as the time between infection and the first appearance of clinical signs. The *prepatent period* refers to the time between infection and when the agent becomes first detectable, and the *period of communicability* is the time during which the infected host is capable of transmitting the agent. The agent for a particular disease can be transmitted via different mechanisms whose identification may allow introduction of specific measures for preventing transmission. *Contact transmission* can occur through *direct* (venereal diseases) or *indirect* contact (excretions, secretions, exhalations). It depends on the survival of the agent in the environment and the extent of contact between infected and susceptible individuals from the host population. *Vehicular transmission* refers to transfer of the agent in inanimate substances (fomite). It requires

prolonged survival of the agent, but allows transfer over long distances and long time periods. Some agents can reproduce during transmission (Salmonella). The presence of *vectors* or *intermediate hosts* can be a requirement for an infectious agent to survive within an eco-system. Under such circumstances, the *definitive host* (usually a vertebrate) allows the agent to undergo a sexual phase of development. In the *intermediate host* (vertebrate, invertebrate), the agent undergoes an asexual phase of development. A *vector* is an invertebrate actively transmitting the infectious agent between infected and susceptible vertebrates through *mechanical* or *biological transmission*. The latter can be *transovarial* allowing maintenance of infection within the vector population or *transtadial* transmission involving transmission between different development stages of vector. *Intrinsic host determinants* include factors such as species, breed, age and sex. The range of *susceptible* host species varies substantially between infectious agents. Many disease agents such as *Mycobacterium bovis* can infect many different animal species. A species is considered a *natural reservoir of infection* if infection can be maintained within the species population without requiring periodic re-introduction. This type of epidemiological scenario can extremely complicate control or eradication of a disease in domestic livestock if the reservoir of infection is a wildlife species. *Host susceptibility* can vary between breeds of a particular animal species such between *Bos indicus* and *Bos taurus* with respect to trypanosomiasis or tick resistance. Variation in *age susceptibility* is probably the most important host variable. Young animals may, for example, be less susceptible to tick-borne diseases than adults. But there can be confounding factors such as immunity in older animals which had been exposed to infection as young animals. Passive resistance in new born animals will result in low incidence of infection in young animals. Susceptibility may vary between sexes due to anatomic and /or physiological differences between sexes such as in the case of mastitis and metritis. A confounding element can be

that one sex may be of higher value to farmers resulting in more care and thereby reduced disease incidence. *Extrinsic determinants* of disease affect the interaction between host and agent. They include factors such as climate, soils and man.

Causation

Most scientific investigations are aimed at identifying cause-effect relationships. Webster's dictionary defines a cause as "something that brings about an effect or a result". A cause of a disease is an event, condition, or characteristic which plays an essential role in producing an occurrence of the disease. Knowledge about cause-and-effect relationships underlies every therapeutic manoeuvre in clinical medicine. The situation is complicated if multiple causes are involved. The *Henle - Koch postulates* developed in 1840 (Henle) and 1884 (Koch) were the first set of criteria used to provide a generally accepted framework for identifying causes of disease. They demanded the following criteria to be met before an agent could be considered the cause of a disease:

- It has to be present in every case of the disease.
- It has to be isolated and grown in pure culture.
- It has to cause specific disease, when inoculated into a susceptible animal and can then be recovered from the animal and identified.

Koch's postulates brought a degree of order and discipline to the study of infectious diseases, but had the following basic assumptions which were often impossible to fulfil. They require that a particular disease has to have only one cause and a particular cause should result in only one disease. The Henle-Koch postulates also have difficulty dealing with multiple etiologic factors, multiple effects of single causes, carrier states, non-agent factors (age, breed) and quantitative causal factors. Based on John Stuart Mill's rules of inductive reasoning from 1856, Evan developed the *unified concept of causation* which is now generally accepted for

identifying cause-effect relationships in modern epidemiology. It includes the following criteria:

- The proportion of individuals with the disease should be higher in those exposed to the putative cause than in those not exposed.
- The exposure to the putative cause should be more common in cases than in those without the disease.
- The number of new cases should be higher in those exposed to the putative cause than in those not exposed, as shown in prospective studies.
- Temporally, the disease should follow exposure to the putative cause.
- There should be a measurable biological spectrum of host responses.
- The host response should be repeatable following exposure to the putative cause.
- The disease should be reproducible experimentally.
- Preventing or modifying the host response should decrease or eliminate the expression of disease.
- Elimination of the putative cause should result in a lower incidence of the disease.
- The relationship should be biologically and epidemiologically plausible

The *web of causation* is often used to describe modern disease problems where presence or absence of disease is not just a matter of the agent being present or absent and the disease occurrence is determined by a complex web of interacting factors involving agent, host and environment. Figure 9 presents the causes of tuberculosis in humans as an example of a web of causation.

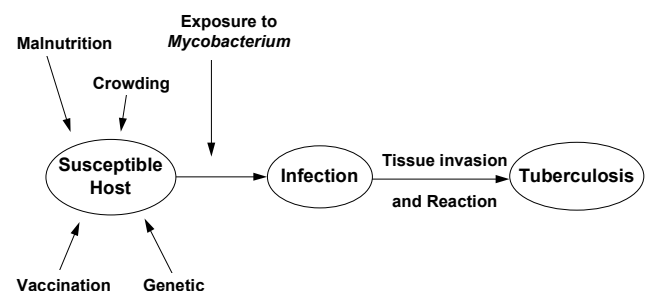


Figure 9: Factors influencing tuberculosis in humans

The term *epidemiological triad* refers to the three components of epidemiological system thinking: *agent*, *host* and *environment*.

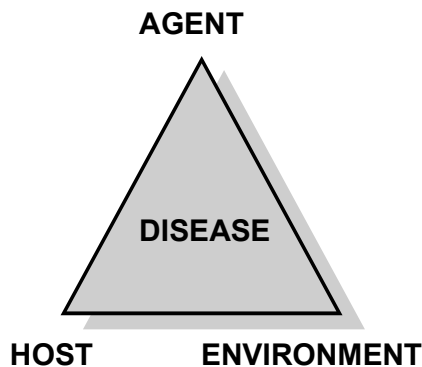


Figure 10: The classic epidemiological triad

Figure 11 shows an example of the different parameters influencing the probability of disease occurrence which are associated with each of the three components of the triad.

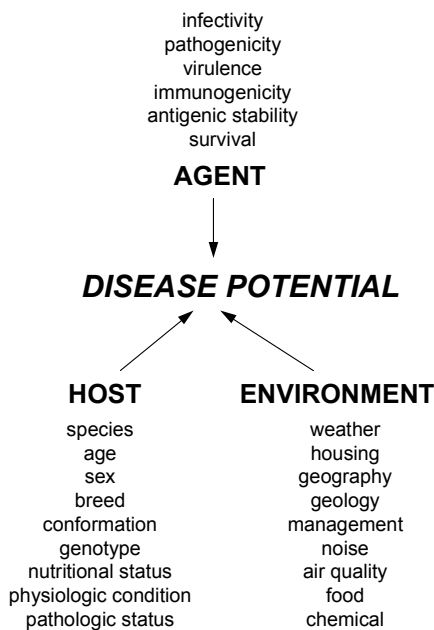


Figure 11: The epidemiological triad

Figure 12 presents a list of most of the factors influencing the occurrence of rhinitis in swine. It illustrates the complexity of the system in which this particular disease does occur. Many of the factors will interact and will have a different effect at varying exposure levels. The Henle/Koch postulates do not provide a suitable mechanism for investigating this type of problem.

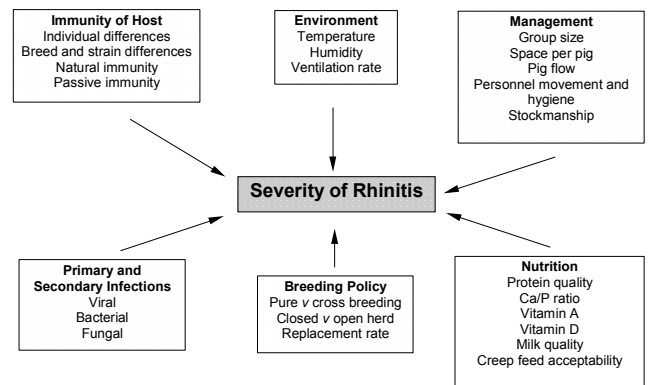


Figure 12: Web of causation for rhinitis in pigs

Causes of diseases can be categorised into *necessary causes* which must be present for a disease to occur (e.g. distemper virus in canine distemper) and *sufficient causes* which are a set of minimal conditions and events inevitably producing disease. In addition, factors can be direct or indirect causes. The strength of a cause as well as the interaction among causes may influence the likelihood of disease occurrence. Figure 13 shows an example of sufficient causes where each of the factor complexes such as weather stress, viruses and *Pasteurella* together represent a sufficient cause for respiratory disease in feedlot cattle.

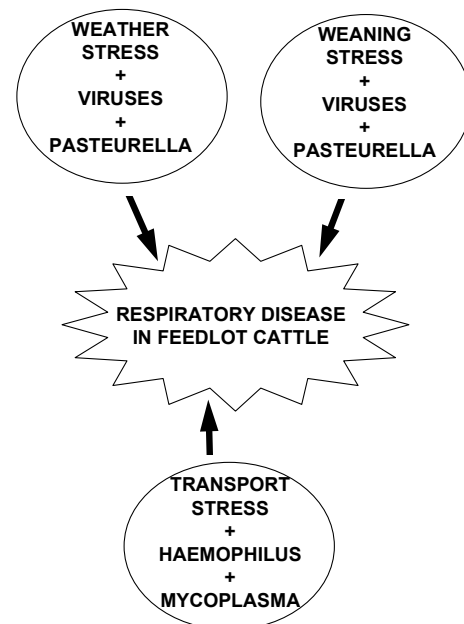


Figure 13: Sufficient and insufficient causes

Figure 14 shows rabies in bat populations as an example of an *indirect cause* of rabies in humans, because infection in foxes may originate in some situations from bat rabies.

The presence of *Pasteurella* is a necessary cause for pasteurellosis, but it is not a necessary cause for pneumonia.

not the true cause. This illustrates that it typically is quite difficult to prove a cause-effect relationship.

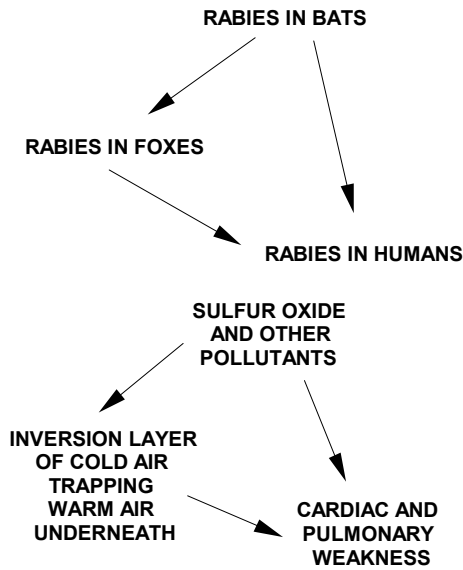


Figure 14: Direct and indirect causes

If one aims at *establishing cause* it is important to realise that it is impossible to prove causal relationships beyond any doubt, but it is possible to use empirical evidence to increase one's conviction of a cause-and-effect relationship to a point where, for all intents and purposes, cause is established. A biological mechanism established in the laboratory under controlled conditions cannot always be assumed to apply under field conditions. The case for causation depends on the strength of the research design used to establish it. In this context, it is important to be aware of the difference between *apparent association* and *true cause*. Figure 15 shows a flow chart of the process leading towards evidence of cause-effect. An apparent association between a potential risk factor and disease status may appear to be present on the basis of, say, a comparison of two proportions. Given this observation the data should be assessed for selection or measurement bias. The likelihood that the observed difference was due to chance variation can be quantified using a statistical test such as the chi-square test. But even if it appears that it is unlikely that the observed difference between the proportions was due to chance, there is still a possibility that the risk factor was a confounding factor and therefore

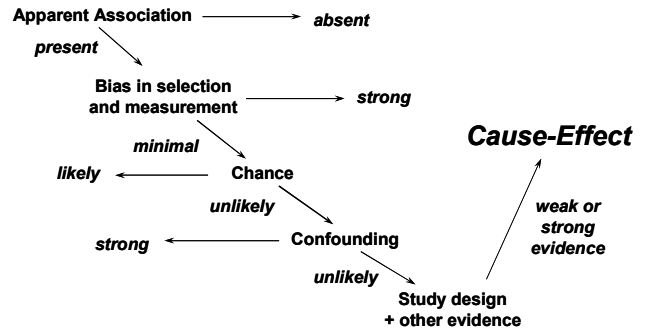


Figure 15: From association to cause-effect relationship

Descriptive Epidemiology

Learning Objectives

At the completion of this topic, you will be able to:

- differentiate between ratios, proportions and rates
- appropriately use prevalence and incidence
- understand the difference between risk and rate as applied to measures of incidence
- understand the meaning of survival probability and hazard rate

Measurement of Disease Frequency and Production

One of the most fundamental tasks in epidemiological research is the quantification of the disease occurrence. This can be done simply on the basis of *counts* of individuals which are infected, diseased, or dead. This information will be useful for estimating workload, cost, or size of facilities to provide health care. More commonly, *counts* are expressed as a *fraction* of the number of animals capable of experiencing infection, disease or death. These types of quantities are used by epidemiologists to express the *probability* of becoming infected, diseased or dying for populations with different numbers of individuals (= populations at risk).

From a mathematical perspective, frequency of disease occurrence can be expressed through *static* or *dynamic* measures. *Static measures* include proportions and ratios. A *proportion* is a fraction in which the numerator is included within the denominator. It is dimensionless, ranging from 0 to 1 and is often expressed as a percentage (x 100). The *ratio* is a fraction in which the numerator is not included in the denominator and it can be with or without dimension. *Dynamic measures* include *rates* which represent the instantaneous change in one quantity per unit change in another quantity (usually time). They are not dimensionless and do not have a finite upper bound. Measures of disease frequency can be based only on new

(=incident) cases of disease or do not differentiate between old and new disease. Figure 16 shows the principles behind incidence measures. They are derived from data for animals which did not have the disease at the beginning of the study period. These animals are followed over time until they develop the disease (eg become lame) or until the observation period finishes (eg sold or end of study period).

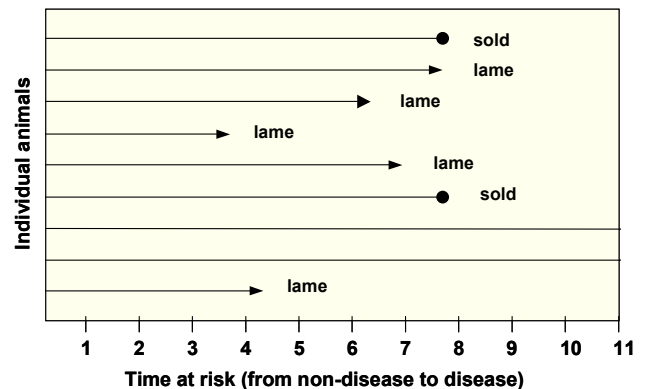


Figure 16: Incidence of disease

Cumulative incidence

The risk of new disease occurrence is quantified using *cumulative incidence*, also called *incidence risk*. It is defined as the proportion of disease-free individuals developing a given disease over a specified time, conditional on that individual's not dying from any other disease during the period. Note that animals have to be disease-free at the beginning of the observation period to be included in the numerator or denominator of this calculation. It is interpreted as an individual's risk of contracting disease within the risk period. The quantity is dimensionless, ranges from 0 to 1 and always requires a period referent. As an example, last year a herd of 121 cattle were tested using the tuberculin test and all tested negative. This year, the same 121 cattle were tested again and 25 tested positive. The cumulative incidence over a period of 12 months would then be calculated as 25/121 which amounts to 0.21. Hence, an individual animal within this herd had a 21% chance of becoming infected over the 12 month period.

Incidence density

Incidence density (also called true incidence rate, hazard rate, force of morbidity or mortality) is defined as the instantaneous potential for change in disease status per unit of time at time t , relative to the size of the disease-free population at time t . The numerator contains the number of new cases over the time period observed and the denominator is the accumulated sum of all individual's time at risk (=population time at risk). This measure does not have an interpretation at the individual animal level. It is expressed in units of 1 per time unit and can exceed 1. Two methods are commonly used for its calculation. One uses an exact denominator calculated as the sum of animal time units during which each animal was at risk, and the other uses an approximate denominator based on the total number of disease free animals at the start of the time period, from which 1/2 of the diseased and 1/2 of the withdrawn animals is being subtracted. As an instantaneous rate it expresses the potential of disease occurrence per unit of time. The following example illustrates the principles behind the calculation of incidence density:

A study was conducted over a period of 12 months to determine the mortality of cows in a village which has a total 100 cows at the beginning of the study.

- 5 cows die after 2 months which means they were $5 * 2 = 10$ animal months at risk
- 2 cows die after 5 months which means they were $2 * 5 = 10$ animal months at risk
- 3 cows die after 8 months which means they were $3 * 8 = 24$ animal months at risk

This means a total of 10 cows die, and these experienced 44 animal months at risk based on the calculation ($5 * 2 + 2 * 5 + 3 * 8$)

- 90 cows survive past the study period which means they were $90 * 12$ months = 1080 animal months at risk

Therefore, the incidence density of cow mortality in this village is calculated as $10 / 1124 = 0.009$ deaths per animal month.

Prevalence

This is the proportion of a population affected by a disease at a given point in time. It can be interpreted as the probability of an individual from the same population having the disease at this point in time. *Period prevalence* refers to disease occurrence over a period of time and *point prevalence* only looks at a single point in time. The prevalent disease cases used in the numerator include old as well as new cases. There is no temporal sequence inherent in the calculation. This means that it is impossible to know when these animals became diseased. If the average duration of disease and cumulative incidence are known, their product can be used to calculate prevalence. As an example of a prevalence calculation, assume a situation where blood samples are taken from a herd of 173 dairy cows to assess the frequency of *Neospora caninum* infection. If 15 of these animals test positive, prevalence can be calculated as $15/173$ amounting to 0.09 (9%). This means that each dairy cow within the herd has a 9% chance of being infected at this point in time.

Comparison of prevalence and incidence measures

Comparing cumulative incidence and prevalence, it is important to realise that only the first includes a temporal sequence.

Cumulative incidence does only include new cases in the numerator, whereas prevalence does not distinguish between old and new cases. Cumulative incidence predicts what will be happening in the future as the probability that similar individuals will develop the condition in the future. This is useful for making decisions about preventive measures such as vaccination. Prevalence describes the probability of having the disease among a group of individuals at a point in time. Every clinician uses this information intuitively during the clinical decision making process. Both measures can be used to make comparisons between risk factors such as when comparing prevalence of disease in vaccinated and non-vaccinated animals. Table 1 presents a comparison of the three methods for expressing disease frequency.

Table 1: Comparison of measures of disease occurrence

	Incidence density	Cumulative incidence	Prevalence
Numerator	New cases occurring during a period of time among a group initially free of disease	New cases occurring during a period of time among a group initially free of disease	All cases counted on a single survey of a group
Denominator	Sum of time periods during which individuals could have developed disease	All susceptible individuals present at the beginning of the period	All individuals examined, including cases and non-cases
Time	For each individual from beginning of follow-up until disease	Duration of period	Single point or a period
How measured	Prospective Cohort study	Prospective Cohort study	Cross-sectional study
Interpretation	Rapidity with which new cases develop over given time period	Risk of developing disease over given time period	Probability of having disease at a particular point in time

An example for the calculation of the different measures of disease occurrence is shown in Figure 17. The calculation is based on a herd of 10 animals which are all disease-free at the beginning of the observation period and are being followed over a period of one year. Disease status is assessed at monthly intervals. Animal A shows up with disease in May and therefore was at risk from January until April. Animal C was withdrawn from the population in August which means that it was at risk of becoming diseased from January to July. A calculation of point prevalence in December would yield an estimate of 50% and in June of 33%. Hence, if the disease process is influenced by seasonal effects and duration of the disease is short, point prevalence estimates will vary substantially depending on when the population at risk was examined. The withdrawals will cause problems when calculating the incidence estimates. For the

Animal	January	February	March	April	May	June	July	August	September	October	November	December	Time at Risk in Months	Diseased	Population	Withdrawal
A					Disease								4	yes	1	no
B													12	no	1	no
C								Withdrawn					7	no	1	yes
D		Disease											1	yes	1	no
E													12	no	1	no
F						Disease							5	yes	1	no
G												Disease	10	yes	1	no
H													12	no	1	no
I													12	no	1	no
J								Withdrawn					5	no	1	yes
Total													80	4	10	2

Point Prevalence in December 0.50 = 4 / 8
 Point Prevalence in June 0.33 = 3 / 9
 Cumulative Incidence per Year 0.50 = 4 / 8
 Incidence Density per Year
 approximate 0.57 = 4 / ((10 - 0.5 * 4 - 0.5 * 2))
 exact 0.60 = 4 / ((4 + 12 + 7 + 1 + 12 + 5 + 10 + 12 + 12 + 5)/12)

Figure 17: Calculation of measures of disease occurrence

cumulative incidence calculation they were excluded, and in the case of incidence density the approximate and the exact calculation resulted in very similar estimates. Any

interpretation of the incidence figures should take the risk period into account, which in this case is one year.

Miscellaneous measures of disease occurrence

Other measures of disease frequency include the *attack rate* which is defined as the number of new cases divided by the initial population at risk. As it is based on the same calculation as cumulative incidence it really is a subtype of cumulative incidence. It is confusing though that despite its name it is in fact a probability and not a rate. The attack rate is used when the period at risk is short.

Mortality rates are applied using a number of different interpretations, and often do not represent a true rate. The *crude mortality rate* has death as the outcome of interest and is calculated analogous to incidence density.

The *cause-specific mortality rate* is estimated for specific causes of death and also calculated analogous to incidence density.

Case fatality rate represents the proportion of animals with a specific disease that die from it. It is a risk measure, not a rate, and is used to describe the impact of epidemics or the severity of acute disease.

Survival

Time to the occurrence of an event such as death or onset of clinical disease can be estimated for many epidemiological data sets describing repeated observations on the same sample of animals. This type of data can be summarised using for example mean survival time. This particular method has the disadvantage that the estimate will depend on the length of the time period over which data was collected. If the interval is too short, survival time is likely to be estimated incorrectly, as only individuals who experienced the event of interest can be included in the calculation. As an alternative, it is possible to calculate an incidence density by using person-years of observation in the denominator and number of events in the numerator. But this calculation assumes that the rate at which the event occurs is constant throughout the period of study. The most appropriate technique for this data is based on

the *survivor* and *hazard function*. The survivor function (=cumulative survival probability) is a functional representation of the proportion of individuals not dying or becoming diseased beyond a given time at risk. It can be interpreted as the probability of remaining alive for a specific length of time. The survival function is often summarised using the *median survival time* which is the time at which 50% of individuals at risk have failed (died or became diseased). The *hazard function* (=instantaneous failure rate, force of mortality, conditional mortality rate, age-specific failure rate) is calculated by dividing the conditional probability of an individual dying or becoming diseased during a specific time interval provided it has not died or become diseased prior to that time divided by the specified *time interval*. This parameter does represent a rate expressing the potential of failing at time t per unit time given survival up until time t . In the context of survival data, *censoring* is an extremely important concept. In the case of *right censoring*, individuals are lost to follow-up or are not dead/diseased at the end of the follow-up period. This particular type of censoring can be easily accounted for in survival analysis by excluding them from the denominators of the calculations following their departure from the population at risk. With *left censoring*, beginning of the time at risk is not known, and the commonly used analysis techniques cannot take account of this type of censoring.

An example calculation for survival data is presented in Figure 18. Animal A survived for 4 months, animal B survived the whole period of 12 months and animal C was removed from the population after 7 months. The number of survivors is based on the actual number of animals still alive after a given time period. The value for the cohort is used as the denominator for calculation of cumulative survival. The number is adjusted for censored observations. The failures represents the number of deaths during a particular time interval. And the hazard rate is the probability of death per unit time (one month in this case). The resulting graphs for

the survivor and hazard functions are shown in Figure 19.

available to perform these calculations are *direct* and *indirect standardisation*.

Animal	Month of year											
	1	2	3	4	5	6	7	8	9	10	11	12
A					Death							
B												
C								CENSORED				
D		Death										
E												
F						Death						
G											Death	
H												
I												
J						CENSORED						
Survivors	10	9	9	9	8	6	6	5	5	5	4	4
Cohort	10	10	10	10	10	9	9	8	8	8	8	8
Survival	1.00	0.90	0.90	0.90	0.80	0.67	0.67	0.63	0.63	0.63	0.50	0.50
Failures	0	1	0	0	1	1	0	0	0	0	1	0
Hazard	0.00	0.11	0.00	0.00	0.13	0.17	0.00	0.00	0.00	0.00	0.25	0.00

Direct standardisation

Direct standardisation involves weighting a set of observed category-specific risk estimates according to a *standard* distribution. First, stratum-specific risk rates are calculated. Then a *standard* population distribution is estimated and the proportion of the standard population in each stratum calculated. The direct adjusted risk rate estimate is obtained as the sum of the products across the strata between

Figure 18: Example calculation for survival data

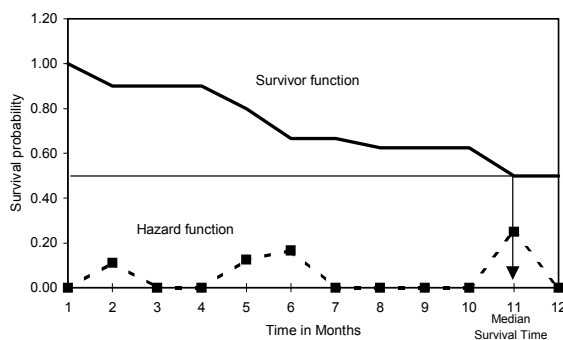


Figure 19: Plots for survivor and hazard functions based on the example calculation

Standardisation of Risk

A crude risk estimate summarises the effects of the specific risk and the subgroup distribution. But in the presence of a confounding factor a crude risk estimate may distort the true pattern. In such a situation, patterns of disease should be described using host-attribute specific risk estimates (e.g. age, sex). For each level of the attribute a separate stratum is formed and stratum-specific risk estimates are calculated. Summary figures can be produced using *standardised* or *adjusted risk* estimates. The two main methods

the proportion of the *standard* population in stratum *i* and the observed risk rate estimate in stratum *i* in the *study* population. As an example, the mortality in humans in 1963 was compared between Sweden and Panama. In this particular year, Sweden had a population size of 7.496.000 and 73.555 deaths resulting in a *mortality rate* of 0.0098 per year. Panama had a population size of 1.075.000 with 7871 deaths giving a *mortality rate* of 0.0073 per year. Based on these figures it appeared that life in Sweden was more risky than in Panama. Figure 20 shows an example of applying the method of *direct standardisation* to this data. It becomes apparent that in his comparison the confounding factor was the difference in age structure between the two populations. Sweden had a much lower mortality in young people, but because it had a large proportion of old people this effect did not come through in the aggregated analysis. After adjustment for differences in age structure it turns out that mortality was higher in Panama with 0.0162 per year than it was in Sweden with 0.015 per year.

		Mortality rate (per year)	
Standard population	Weight	SWEDEN	Product between stratum-specific rate and weight
0-29	0.35	0.0001	0.00004
30-59	0.35	0.0036	0.0013
60+	0.30	0.0457	0.0137
Standardised rate		0.0150	
Crude rate		0.0098	
		PANAMA	
0-29	0.35	0.0053	0.0019
30-59	0.35	0.0052	0.0018
60+	0.30	0.0416	0.0125
Standardised rate		0.0162	
Crude rate		0.0073	
Rate ratio (Sweden / Panama) based on			
Standardised rates		0.9290	
Crude rates		1.3425	

Figure 20: Example of direct standardisation calculations

Indirect standardisation

An alternative approach to standardisation of risk estimates is called *indirect standardisation*. In this case, a *standard population* does not supply the weighting distribution, but a set of stratum-specific risk estimates which are then weighted to the distribution of the *study population*. This technique is used if stratum-specific risk estimates are not available. But in order to be able to use the method, stratum-specific risk estimates for the *standard population* and the frequency of the adjusting factor in the *study population* have to be available. As a first step, the expected number of cases is calculated on the basis of the sum of the products between stratum-specific rates for the *standard population* and the total number of individuals in each stratum in the *study population*. The standardised morbidity or mortality ratio (SMR) is calculated using the number of observed cases divided by the number of expected cases, and the indirect adjusted risk is obtained from multiplying the overall observed risk in the *study population* with the SMR.

Figure 21 shows an example of an indirect standardisation where there are two different levels of exposure to a risk factor. The *standard* and the *study* populations are stratified into young and old people. For the standard population, the overall crude rate of

disease is known to be 0.0008 and the one for the two age groups is 0.00005 and 0.002. These figures are used to estimate the expected number of cases for each age and exposure category. With exposure 1 in young people in the standard population 5 cases would be expected, but 50 did actually occur. For each exposure, these observed and expected values are then added up separately. Within the each exposure group a standardised mortality ratio is estimated as the ratio of total observed and total expected resulting in an SMR of 7.71 for exposure group 1. An indirect adjusted risk rate can be calculated as the product of the crude rate in the *standard population* and the SMR. The resulting figures are very different from the crude estimates for each exposure level, which had been biased by the difference in age structure between the two exposure groups.

Age Group	Standard population	Exposure 1		Exposure 2	
		observed	expected	observed	expected
Young					
Cases		50	5	5	0.5
Person-Years		10000		1000	
Incidence Rate	0.0005	0.005		0.005	
Old					
Cases		4	2	40	20
Person-Years		1000		10000	
Incidence Rate	0.002	0.004		0.004	
Crude Rate	0.0008	0.0049		0.0041	
Total Cases	0	54	7	45	20.5
Total Population	0	11000		11000	
Calculation		0.0008*7.71	54 / 7	0.0008*2.2	45 / 20.5
Standardized Mortality ratio			7.71		2.20
Indirect adjusted rate		0.0062		0.0018	

Figure 21: Example calculation for indirect standardisation

Analytical Epidemiology

Learning Objectives

At the completion of this topic, you will be able to:

- design an observational or non-observational epidemiological study and understand their respective differences as well as advantages/disadvantages
- determine the quantitative contribution of a risk factor to a disease causation using one or more of the following procedures: Relative risk; odds ratio; attributable risk, attributable fraction
- understand the basic concepts of statistical hypothesis testing
- recognise the potential sources of bias in observational studies
- understand the concepts of confounding and interaction

Introduction

The major aims of epidemiology are to *describe* the health status of populations, to *explain* the aetiology of diseases, to *predict* disease occurrence and to *control* the distribution of disease. An understanding of causal relationships is the basis of the last three objectives. Such associations between causes and disease occurrence can be discovered through individual *case studies*, by experimental *laboratory studies* and by *field studies*. *Case studies* focusing on individual sick animals have long been at the centre of clinical knowledge. They are based on direct personal observations relating to anatomical structure and physiological function, which can be quantified and are systematic but still largely qualitative. While these observations can be extremely intensive and detailed their disadvantage is their subjectivity and the possibly extreme variation between cases. In the *laboratory experiment* - the classic experiment - great precision in measurements and optimal control of influencing variables can be achieved resulting in sound inferences. The disadvantage is that it is usually not possible to represent the myriad of factors affecting disease occurrence in the natural environment of the animal and it may be

difficult to work with sufficient numbers of animals to represent true variation between animals in the natural population. A *field study* is conducted in the natural environment of the animals and measurements are made on sick as well as healthy animals. The differences between sick and healthy animals can be described with respect to the frequency of presence or absence of potential risk factors. With this type of study, animals are exposed to all the known and unknown environmental factors present in their natural environment.

Field research is *empirical* and involves measurement of variables, estimation of population parameters and statistical testing of hypotheses. It is of a *probabilistic nature* in that as a result of a population study it will not be possible to predict with certainty which animal will develop a particular disease given the presence of certain risk factors. But it will be possible to predict how many cases of the disease will occur in the population in the future. Field research involves comparisons among groups in order to estimate the magnitude of an *association* between a putatively causal factor and a disease. The objective is to assess if there is the potential of a cause-effect relationship between a single or multiple risk factors and the disease.

The interrelatedness of phenomena within a biological system complicates the situation for an investigator who will always have to select a segment of the system for the investigation. Attempting to isolate the segment from the rest of the system can result in an outcome which does not represent the real situation in the system anymore.

Analytical epidemiology is aimed at determining the *strength, importance* and *statistical significance* of epidemiological associations. The process typically begins with data collection and eventually leads to data analysis and interpretation. The data collection can be based on a *survey* or a *study*. Both terms are often used interchangeably. A *survey* typically involves *counting* members of an aggregate of units and *measuring* their characteristics. In contrast, a *study* is aimed at

comparison of different groups and investigation of cause-effect relationships. Both designs can be based on a *census* where all members of the population are included thus allowing exact measurement of variables of interest, or alternatively on a *sample* where a subset of the population is included, thereby providing only estimates of the variables of interest.

Epidemiological Studies

Epidemiological studies are broadly categorised into *non-observational* (or *experimental studies*) and *observational studies*. The first group includes clinical trials or intervention studies and the basic principle is that the design of the study involves deliberately changing population parameters and assessing the effect. With this type of study an attempt is made to simplify observation by *creating* suitable conditions for the study. The second group assumes that the study does not interfere with the population characteristics. Here, the investigator is only allowed to *select* suitable conditions for the study. *Observational studies* can be further categorised into *prospective cohort studies* and *retrospective studies* as well as *cross-sectional studies*. *Retrospective studies* include case-control and retrospective cohort studies. Another type of observational study is the *longitudinal study* which is a mix between prospective cohort and repeated cross-sectional studies. Within the group of observational studies mixtures of study designs are common, such as for example the case-cohort study. The *case series* is a separate group of studies and frequently used in a clinical context.

Non-observational studies

This type of study typically involves dividing a group of animals into a subgroup which is being *treated* and another subgroup which is being left *untreated* and acts as a control (see Figure 22). The decision to treat an animal or leave it untreated is typically based on random allocation - *randomisation*. After a period of time the status with respect to a response variable (e.g. disease status) is assessed for each animal. Summary measures of the

response are then compared between both subgroups. Differences in the summary values suggest the presence of an effect of the *treatment* on the response variable. Non-observational or experimental studies can be conducted as *laboratory experiments* or as *field studies* such as *clinical trials*. The latter are usually used to evaluate therapeutic or preventive effects of particular interventions, but are also useful to investigate etiologic relationships. The *non-observational study* provides the researcher with effective control over the study situation. If the sample size is large enough a well-designed experiment will limit the effect of unwanted factors even if they are not measurable. Control of factors other than the treatment which are likely to have an effect on disease can be achieved by using them to define homogeneous subgroups with respect to the status of these variables - *blocking* or *matching*- within which treatment is then allocated randomly. The possibility to have excessive control over the study situation can become a weakness of the non-observational approach as it may not be representative of the real situation in the biological system anymore. *Clinical trials* are considered the method of choice for investigation of causal hypotheses about the effectiveness of preventive measures, and compared with the other types of field studies they can provide the strongest evidence about causality. There is less opportunity for systematic error compared with the observational studies. Amongst their disadvantages are the following characteristics. They require large groups, are costly, bias may be introduced through selection error and the required duration can be long if disease incidence is low.

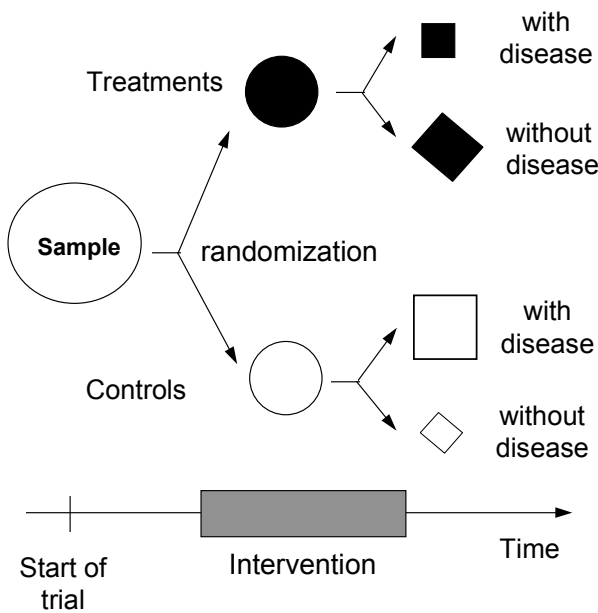


Figure 22: Schematic structure of an experimental study

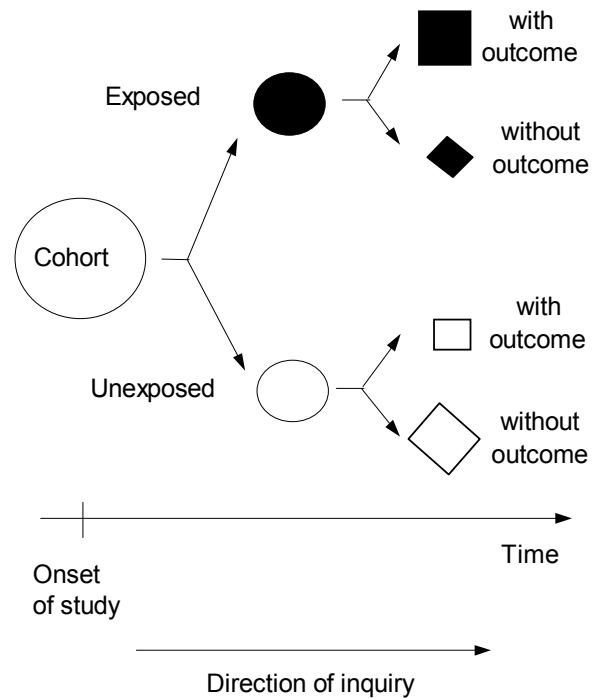
Observational field studies

In epidemiological research, the *observational field study* is one of the most frequently used techniques. This group includes three major study designs, the *prospective cohort*, the *case-control* and the *cross-sectional study*.

Prospective cohort study

The *prospective cohort study* is based on selecting two groups of non-diseased animals, one exposed to a factor postulated to cause a disease and another one unexposed to the factor (see Figure 23). They are followed over time and their change in disease status is recorded during the study period. The prospective cohort study is the most effective observational study for investigation of causal hypotheses with respect to disease occurrence. It provides disease incidence estimates which are more meaningful than prevalence data for establishing cause-effect relationships. Cohort studies can be used to study rare exposures and it is possible to minimise bias. But the investigator has to keep in mind that given its observational nature the prospective cohort study does not provide proof of causality, it can only demonstrate temporality. Prospective cohort studies often require a long duration which increases the potential for confounding effects and therefore affects the ability to demonstrate causality. In the case of rare disease large groups are necessary. Losses to

follow-up can become an important problem and cohort studies are often quite expensive.

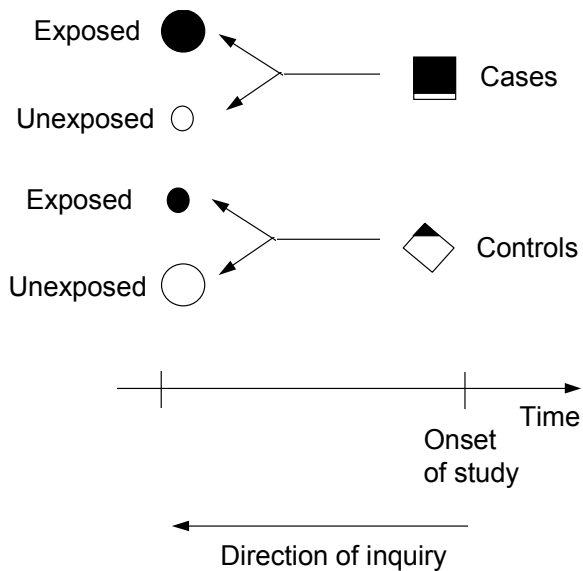


Q: What will happen?

Figure 23: Schematic diagram for prospective cohort study

Case-control study

In a *case-control study* animals with the disease (*cases*) and without the disease (*controls*) are selected (see Figure 24). Their status with regard to potential risk factors is then examined. This type of design can be used effectively for the study of low incidence diseases as well as of conditions developing over a long time. Case-control studies allow the investigation of preliminary causal hypotheses and are quick and of relatively low cost. Their disadvantages include that they cannot provide information on the disease frequency in a population. Furthermore, they are not suitable for the study of rare exposures, and data collection is reliant on the quality of past records. It can also be very difficult to ensure an unbiased selection of the control group. The representativeness of the sample selection process is difficult to guarantee. This problem applies typically to the selection of the control group.

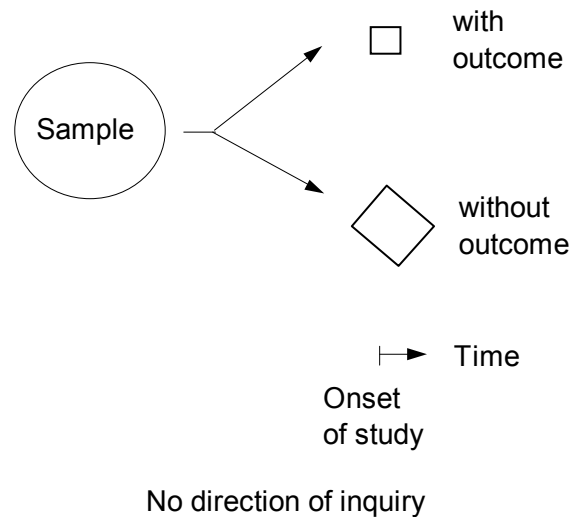


Q: What happened?

Figure 24: Schematic diagram for case-control study

Cross-sectional study

In a *cross-sectional study* a random sample of individuals from a population is taken at one point in time. Individual animals included in the sample are examined for the presence of disease and their status with regard to other risk factors (see Figure 25). This type of study is useful for describing the situation at the time of data collection and it allows determining prevalence. The data should be based on a representative sample drawn from the population. Cross-sectional studies are relatively quick to conduct and cost is moderate. Disadvantages include that they provide only a "snapshot in time" of the disease occurrence. It is difficult to investigate cause-effect relationships and difficult to obtain sufficiently large response rates which will adversely affect the representativeness of the sample. Any inference from this type of study has to take into account the potential for confounding relationships between risk factors.



Q: What is happening?

Figure 25: Schematic diagram of cross-sectional study

Comparison of the three basic observational study designs

The characteristics of the three main observational field study designs are compared in Figure 26. In summary, the *cohort study* is the design amongst *observational studies* which provides the best evidence for the presence of *cause-effect relationships*, because any putative cause has to be present before disease occurs. But as it is based on pure observation within a largely uncontrolled environment it is possible that there are still other unmeasured (=confounding) factors which have produced the apparent cause-effect relationship. The *cohort study* is inefficient for studying rare diseases, which in turn is a particular strength of the *case-control study*. A carefully designed *cross-sectional study* is more likely to be representative of the population under study than a *case-control study*. New etiologic hypotheses can be developed efficiently using *cross-sectional studies*, but less so with *cohort studies*.

With any scientific investigation an awareness of the limitations and advantages of particular study designs is essential during the planning, analysis and interpretation phases of epidemiological studies. Experimentation and determination of biological mechanisms provide the most direct evidence of a causal

relationship between a factor and a disease. Epidemiological field studies can provide strong support for causal hypotheses. Combined epidemiological and other evidence can lead to the conclusion that a causal hypothesis becomes highly probable.

Criteria	Cross-sectional study	Case-control study	Prospective cohort study
Sampling	random sample of study population	separate samples of diseased and non-diseased units	separate samples of exposed and non-exposed units
Time	one point	usually retrospective	follow-up over specified period
Causality	association between disease and risk factor	preliminary causal hypothesis	causality through evidence of temporality
Risk	prevalence	none	incidence density, cumulative incidence
Comparison of risks	relative risk, odds ratio	odds ratio	relative risk, odds ratio

Figure 26: Comparison of observational field studies

Concept of Risk

Any investigation into the cause-effect relationships between potential risk factors and an outcome parameter such as disease or death involves calculation of *risks*. A generic definition of risk says that it is the probability of an untoward event. *Risk factors* include any factors associated with an increased risk of becoming diseased or to die. *Exposure* to a risk factor means that an individual has, before becoming ill, been in contact with the risk factor. *Risk assessment* is performed intuitively by everyone on a daily basis. Most of the time it is done based on *personal experience*, but this approach is *insufficient* to establish a relationship between exposure and disease particularly with an infectious process involving long latency periods, with exposure to the risk factor being common, with diseases of low incidence or of high prevalence, or in the presence of multiple exposures. Under any such circumstances it is preferable to base a comparison on quantitative estimates of risk such as *cumulative incidence*.

The relationship between measures of disease frequency and risk factors can be used for

predictive purposes where knowledge of the disease risk in individuals with the risk factor present is used to manage disease. For *diagnostic* purposes, the presence of a known risk factor in an individual increases the likelihood that the disease is present. If it is a strong risk factor, absence can be used to *rule out* specific diseases. If the risk factor is also the cause of the disease, its removal can be used to *prevent* disease. When assessing the cause-effect relationship, one should always be aware of potential *confounding* factors.

Identification of Risk Factors

Epidemiological studies are conducted to identify risk factors through the comparison of *incidence* or *prevalence* between groups exposed and not exposed to a risk factor. Probabilities of disease occurrence can be compared using measures of *strength of association* or measures of *potential impact*. The first group involves calculation of ratios such as *relative risk* and *odds ratio* which measure the magnitude of a statistically significant association between risk factor and disease. They are used to identify risk factors, but do not provide information on absolute risk. In contrast, measures of *potential impact* include differences such as the *attributable risk* or fractions such as the *attributable fraction*. These allow quantifying the consequences from exposure to a risk factor, and are used to predict, quantify the effect of prevention and to plan control programs.

Relative risk

The *relative risk* (= RR; risk ratio, cumulative incidence ratio or prevalence ratio) is used if the following question is asked: How many times more (or less) likely are exposed individuals to get the disease relative to non-exposed individuals? It is calculated as the ratio of *cumulative incidence* or *prevalence* between exposed and non-exposed individuals. *Cumulative incidence ratio* and *prevalence ratio* are similar if disease duration is unrelated to the risk factor. The RR is interpreted as follows: The disease is RR times more likely to occur among those exposed to the suspected risk factor than among those with no such exposure. If RR is

close to 1, the exposure is probably not associated with the risk of disease. If RR is greater or smaller than 1, the exposure is likely to be associated with the risk of disease, and the greater the departure from 1 the stronger the association. RR cannot be estimated in *case-control studies*, as these studies do not allow calculation of risks.

Odds ratio

The *odds ratio* (= OR; relative odds, cross-product ratio and approximate relative risk) is calculated as the ratio between the odds of disease in exposed individuals and the odds of disease in non-exposed individuals. It is interpreted as the odds of having the disease among those exposed to the suspected risk factor being OR times the odds of disease among those with no such exposure. If OR is close to 1, the exposure is unlikely to be associated with the risk of disease. For an OR greater or smaller than 1, the likelihood that the exposure is associated with risk of disease increases, and the greater the departure from 1 the stronger the potential cause-effect relationship. In contrast to RR, OR can be used irrespective of the study design, including case-control studies. OR is also insensitive to whether death or survival is being analysed. OR can be used to estimate RR if the disease is rare (less than 10%). Odds and risks have the same numerator (=the number of diseased), but differ in the denominator, which in the case of odds includes only events which are not in the numerator and in the case of risks includes all events.

Rate ratio

If the researcher asks the question "How much more likely it is to get cases of disease in the exposed compared with the non-exposed population?", the *rate ratio* (incidence rate ratio) is the parameter of choice. It is calculated as the ratio of *incidence density* estimates in exposed and unexposed individuals. Similar to RR and OR, if its value is close to 1, it is unlikely that the exposure is associated with the disease frequency. The further the value from unity, the more likely it is that the exposure is related to disease frequency. This quantity can only be

estimated on the basis of data from cohort studies.

Attributable risk

The question "What is the additional risk of disease following exposure, over and above that experienced by individuals who are not exposed?" can be answered through calculation of the *attributable risk* (=AR, risk difference, excess risk, cumulative incidence difference or prevalence difference). AR is estimated through subtracting *cumulative incidence* or *prevalence* of disease in non-exposed from the corresponding values in exposed individuals. It makes the assumption that the risk of disease in the un-exposed group represents the background risk of disease. The AR is interpreted as the risk of developing the disease being increased by AR for those individuals exposed to the risk factor. Different estimates are obtained for AR in the *exposed group* and AR in the *population* (PAR). PAR can be estimated by multiplying AR with the prevalence of the risk factor in the population. The information contained in AR combines the *relative risk* and the *risk factor prevalence*. The larger the AR, the greater the effect of the risk factor on the exposed group. The parameter cannot be estimated for most case-control studies.

Number needed to treat

Evidence-based medicine has resulted in the definition of a new method for expressing attributable risks which is considered to be easier to understand in a clinical context. It is called the *number needed to treat* (NNT), and is interpreted as the number of animal patients needed to treat with a therapy during the duration of a trial in order to prevent one bad outcome. It is calculated as the inverse of the attributable risk (i.e. $1 / AR$). When applying this parameter, it is important to take the follow-up period into account which was used to generate it. The number needed to harm can also be calculated.

Attributable fraction

The *attributable fraction* (= AF; etiologic fraction, attributable risk) is used to answer the question: "What proportion of disease in the exposed individuals is due to the

exposure?". AF is calculated as the proportion that the attributable risk represents within total disease risk in exposed individuals. Different estimates are calculated for AF in the *exposed group* and AF in the *population* (PAF). PAF can be estimated through dividing PAR by the disease prevalence in the population. It is interpreted as the probability that randomly selected individuals from a group/population develop the disease as a result of the risk factor. If the proportion exposed declines in the general population, PAF also decreases, even if RR remains the same. A high PAF implies that the risk factor is important for the general animal population. AF cannot be estimated for most case-control studies.

	Disease	No disease	Total
Exposed	a	b	a + b
Non-exposed	c	d	c + d
Total	a + c	b + d	N

$$RR = \frac{a}{a + b} \bigg/ \frac{c}{c + d}$$

$$OR = \frac{a}{b} \bigg/ \frac{c}{d} = \frac{a \times d}{c \times b}$$

$$AR = \frac{a}{a + b} - \frac{c}{c + d}$$

$$AF = \frac{AR}{a / (a + b)}$$

$$NNT = \frac{1}{AR}$$

Figure 27: Calculation of different measures for comparing risk factors

Vaccine efficacy

Vaccine efficacy (= VE, prevented fraction) stands for the proportion of disease prevented by the vaccine in vaccinated animals. VE is estimated through subtracting cumulative incidence in vaccinated animals from cumulative incidence in unvaccinated animals, and dividing the resulting value by the cumulative incidence in unvaccinated animals.

Calculation of measures for comparing risk factors

The recommended method for calculating the different quantities is to first set up a 2-by-2 table as shown in Figure 27. Many computer programs will automatically perform the required calculations and also include estimates of confidence intervals.

Example calculation for comparison of risk factors

Data on piglet mortality and MMA occurrence has been collected on a piggery

with two farrowing sheds and a total of 200 sows with equal numbers going through each shed. The design of one of the two sheds allows for easy disinfection and cleaning (=good hygiene), whereas the other shed is very difficult to clean (=poor hygiene). The relevance of the different epidemiological measures can be illustrated by estimating the effect of shed hygiene as a potential risk factor affecting the cumulative incidence of piglet mortality (measured as occurring or not occurring on a litter basis) and mastitis-metritis-agalactiae complex (MMA) in sows. Summary data for 200 sows and their litters over a period of 6 months provides the information listed in Table 2. The figure presented in the table indicate that the risk factor hygiene status of the shed has the same strength of association (RR= 5) with both cumulative incidence of piglet deaths and MMA. Attributable risk is considerably higher for litter deaths, because they are more common. Hence, the probability of having piglet deaths in a litter given the presence of the risk factor is much higher than of having a sow with MMA. Control of the risk factor (improving the hygiene standard of the farrowing shed) is clearly justified on the basis of the economic benefits resulting from decreasing piglet mortality, but not necessarily, if it were only to control the incidence of MMA alone. The proportion of cases (litters with piglet deaths or sows with MMA) due to the presence of the risk factor (the attributable fraction) is in both cases the same.

Table 2: Example calculation for a risk factor comparison

		Cumulative Incidence	
Hygiene Status	Number of Sows	Litters with Deaths	MMA
Poor	100	0.25	0.05
Good	100	0.05	0.01

Epidemiological Measures		
Relative Risk	5	5
Attributable Risk	0.20	0.04
Attributable Fraction	0.8	0.8

From Association to Inference in Epidemiological Studies

Investigating the *relationships* between potential risk factors (such as breed of an animal) and the outcome variable of interest (such as infection status of an animal) requires an evaluation of an *observed difference* such as between the prevalence estimates of infection for each of the breed categories. The objective in this case could be to find out if the *probability* of infection of an individual animal is *dependent* on its category of breed. In other words the question to be asked would be “Is the risk of infection for an individual animal any different if the animal belongs to breed A or breed B?”. If there is *dependence* between breed and infection status, a comparison of the two variables (infection status and breed) using the data collected during the study should show a difference between the proportion of diseased in animals of breed A and the proportion of diseased in animals of breed B which is unlikely to have occurred by *chance*. Statistical methods are used to quantify the probability that the observed difference is the result of chance variation. In this example, a chi-square test could be used to test the relationship between the two variables for *statistical significance*. If the chi-square value was larger than 3.84, the associated p-value would be less than 0.05. This means that assuming that there is no real difference the observed difference between the two proportions would be expected to occur just due to chance variation less than 5 times out of 100 similar samples taken from the study population. It can be concluded from this result that the two variables are *statistically significantly associated*. It is important to remember though, that this result is not sufficient to prove that there is a cause-effect relationship between breed category and disease status.

In the following example, it is assumed that the proportion of diseased animals is 0.30 in breed A and 0.50 in breed B animals. Using a 2-by-2 table, a chi-square value of 8.33 with 1 degree of freedom and the associated p-value of 0.004 can be calculated (see Table 3). This

p-value indicates that the observed difference between the two proportions would be expected to occur due to chance variation alone less than 4 times in 1000 similar samples from a study population assumed not to have a difference between breeds. The result of this statistical analysis therefore allows the conclusion that the risk of infection in this population is not independent from breed and that the two variables are statistically significantly associated. Hence, animals of breed A are less likely to become infected than animals of breed B.

Table 3: Comparison of risk of infection and breed

Breed	Infection status		Prevalence (proportion)
	positive	negative	
A	30	70	0.30
B	50	50	0.50

More formally *chance variation* (= random error) can result in *Type I error* (= α - error) or false positive) or *Type II error* (= β - error or false negative). As explained above, the *p-value* is the probability that the observed difference could have happened by chance alone, assuming that there is no difference (likelihood of an α - error). *Statistical power*, on the other hand, stands for the probability that a study will find a statistical difference if it does in fact exist (1 - likelihood of a β error). See Figure 28 for a schematic representation of the relationship between the different types of random error and hypothesis testing.

In addition to chance error, any study can potentially be affected by *bias* (= systematic error). This type of error can be caused by any *systematic* (non-random) *error* in design, conduct or analysis of a study resulting in a mistaken estimate of an exposure's effect on the risk of disease. For example, *selection bias* refers to differences between the study and target population. *Misclassification* or *measurement error* is commonly the result of errors in the classification of disease status.

		True Situation	
		Difference exists (H_1)	No difference (H_0)
Conclusion from Hypothesis Test	Difference exists (Reject H_0)	* (Power or $1 - \beta$)	I (Type I error or α error)
	No difference (Reject H_1)	II (Type II error or β error)	*

Figure 28: Correct Decisions and Errors in Statistical Hypothesis Testing

Confounding

But even if chance or systematic error has been minimised, any observed association can still potentially be the consequence of a *confounding* factor. This particular effect refers to a situation where an independent risk factor is associated with the disease as well as another risk factor, and thereby may wholly or partially account for an apparent association between an exposure and disease. Stratified data analyses can be used to test for the presence of confounding. This means that the association between exposure and outcome is assessed in separate analyses for each level of the hypothesised confounding factor (=controlling or adjusting for the confounding factor). If the strength of the association between exposure and outcome weakens after adjusting for the confounder, meaning the relative risks for each of the strata are close to unity, then there is a potential risk for confounding.

As an example for such a confounding relationship, during the analysis of data from a study of leptospirosis in dairy farm workers in New Zealand the investigators discovered that wearing an apron during milking was apparently associated with an increased risk of contracting leptospirosis. Naïve interpretation of the data could therefore have resulted in the conclusion that if dairy farm workers wanted to reduce the risk of leptospirosis infection they should not wear an apron during milking. But before publicising this result, the investigators found that the risk of infection seemed to increase with herd size, and more importantly farmers with larger herds were found to be more likely to wear aprons during milking than farmers

with smaller herds (see Figure 29). The authors concluded that the apparent association between wearing an apron and leptospirosis infection was in reality a confounding effect of the true effect of herd size.

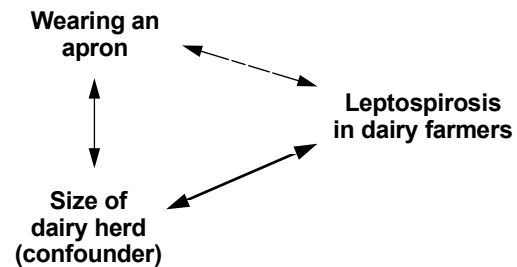


Figure 29: Example of confounding relationship

Interaction

In most biological systems, multiple factors will influence the risk of disease occurrence. Any estimation of effects becomes more difficult if these factors are not independent from each other, meaning the effect of one factor does depend on the level of another. This relationship is called *interaction*. It reflects a biological property of the joint effect of these factors and can manifest itself as either *synergism* or *antagonism*. Interaction is considered to be present, when the combined effects of 2 variables differ from the sum of the individual effects at a defined scale. If there is no interaction, stratum specific relative risks or odds ratios should be equal. Figure 30a shows an example of two factors which do not interact. This becomes evident after stratifying on factor 1 (=holding this factor constant), as the stratum-specific relative risk estimates are both 2. An example for interaction between two risk factors is shown in Figure 30b. Here, after stratifying on GnRH treatment the stratum-specific relative risk estimates quantifying the effect of prostaglandin treatment on incidence of pre-breeding anestrus in cows vary substantially between the strata.

a: no interaction

Factor A	Factor B	Cumulative Incidence	Relative Risk (within stratum)	Relative Risk (compared with both factors being absent)
present	present	0.16	2	8
	absent	0.08		4
absent	present	0.04	2	2
	absent	0.02		reference

equal RR's

b: with interaction

GnRH	Prostaglandin	Incidence of Pre-breeding Anestrus	Relative Risk (within stratum)	Relative Risk (compared with both factors being absent)
present	present	0.13	0.37	1.3
	absent	0.29		2.85
absent	present	0.13	1.35	1.31
	absent	0.1		reference

different RR's

Figure 30: Examples of interaction and no interaction in statistical relationships

The Venn diagram is a useful method for visually presenting interaction relationships between a range of risk factors. Figure 31 presents an example of such a diagram based on an epidemiological field study of feline urological syndrome (FUS) in male cats. The diagram indicates that the relative risk for FUS for male cats which are castrated and have been fed high levels of dry cat food is 33.6, whereas it is only 5 and 4.36 for castration and feeding high levels of dry cat food alone, respectively. This suggests that there might be an interaction between the two risk factors.

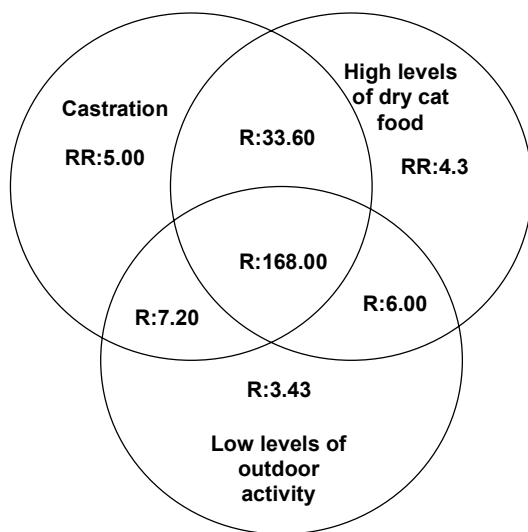


Figure 31: Venn diagram relationships between 3 risk factors for feline urological syndrome in male cats

Sampling of Animal Populations

Learning Objectives

At the completion of this topic, you will be able to:

- to identify in an example, define and differentiate the terms related to sampling methodology
- to give advantages/disadvantages of each sampling method
- to select the appropriate sampling strategy for a particular situation

Introduction

The main objectives of sampling are to provide data which will allow making inferences in relation to a larger population on the basis of examining a sample in relation to, for example, presence/absence of animal disease or other parameters of interest. Inferences might relate to proving that disease is not present, to detecting presence of disease or establishing the level of disease occurrence. The objective could also be to describe levels of milk production in a population of dairy cattle or more generally provide a descriptive analysis of an animal production system for example in an African country.

Data sources

Any *data sources* for epidemiological analyses have to be evaluated with respect to their *completeness*, *validity* and *representativeness*. The data can be collected as part of *routine data collection* which includes laboratory submissions, disease surveillance programmes, industry- or farm/bureau based data recording systems and abattoirs. More recently, *structured data collection* has been found to provide a more effective way for regular monitoring of disease / production. And finally, data can be collected as part of *epidemiological studies*.

Data which is based on *laboratory submissions* is useful for detecting disease. It can become the basis of case series and case-control studies. It does not provide sufficient data to allow prevalence estimation, because

the numerator and denominator are likely to both be biased. In isolation, laboratory submissions do not provide information about causation !!!! They are also not useful for evaluation of therapies or economic effects.

The data collection process can include the whole population of interest (= *census*) or it can be restricted to a *sample*. The latter has the advantage over the *census* that results can be obtained more quickly. A sample is less expensive to collect, and sample results may be more accurate as it is possible to make more efficient use of resources. In addition, *probability samples* result in probability estimates which allow inferences to be used for other populations. *Heterogeneity* in the results can be reduced by targeted sampling of particular sub-groups within the population. Involvement of the whole population such as is necessary for a census may not be possible due to logistic or administrative problems, so that sampling becomes the method of choice.

The sampling process (see Figure 32) can be described using the following terminology. The *target population* represents the population at risk. The population effectively sampled is called the *study population*. Frequently, the target population is not completely accessible, so that it differs to a possibly unknown extent from the study population. It is then necessary to use common sense judgement in order to assess the representativeness of the study population in relation to the target population. The *sampling frame* lists all sampling units in the study population, and is an essential requirement for probability sampling. *Sampling units* are the individual members of the sampling frame. The *sampling fraction* is calculated as the ratio between sample size and study population.

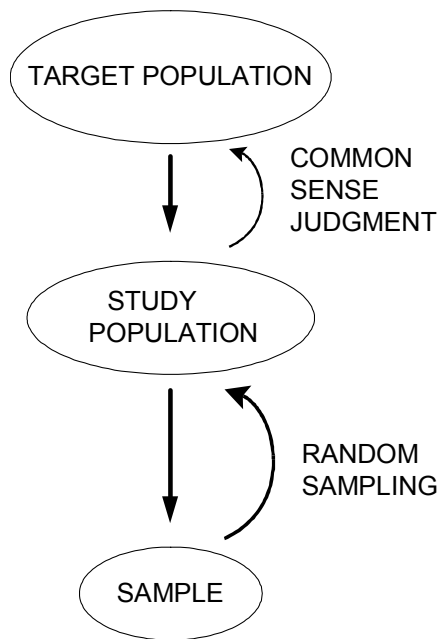


Figure 32: The sampling process

The aim of the sampling process is to draw a sample which is a *true representation* of the population and which leads to estimates of population characteristics having an acceptable *precision* or *accuracy*. Samples can be selected as probability or non-probability samples. With *non-probability sampling*, the investigator chooses the sample as a *convenience sample*, where the most easily obtainable observations are taken or as a *purposive* or *judgmental sample* where deliberate subjective choice is used in deciding what the investigator regards to be a 'representative sample'. The main disadvantage of the non-probability sampling approach is that 'representativeness' cannot be quantified. *Probability sampling* requires random selection of the sample. Sampling units will be accessed through *simple random sampling* where each animal in the study population has the same probability of being selected, independently from any other animal, or through *systematic sampling* where the first unit is selected randomly followed by selection at equal intervals.

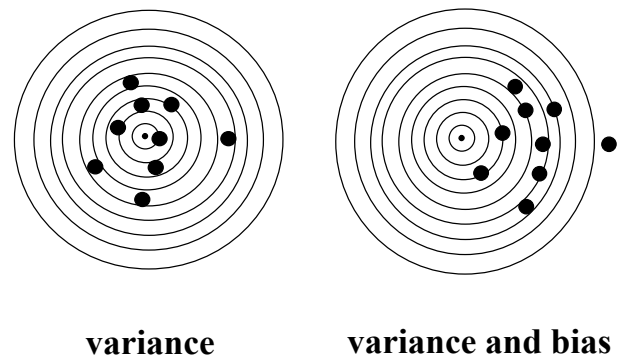


Figure 33: Variance and bias

The aim of probability sampling is to obtain estimates of a variable of interest which are as close as possible to the true (albeit unknown) value for the target population. The result should be unbiased and the effect of sampling variation should be minimal. The probability sample will be subject to sampling and systematic error. *Sampling error* can be quantified and expressed through estimates of variance or confidence limits. *Systematic error* or *bias* can manifest itself as *non-observational error* (selection bias) such as through non-inclusion or non-response, and through *observational errors* including response error and measurement error.

Confidence intervals are now commonly used to measure sampling variability (not bias). It expresses how far away a sample estimate can be from the true value. The correct interpretation of a 95% confidence interval is that given repeated sampling and calculation of 95% confidence intervals for each sample estimate, 95% of them will include the true estimate. Variance or the width of confidence intervals can be influenced through sample size, the selection procedure or mathematical methods. Doubling the sample size will half the variance and quadrupling the sample size halves the confidence interval. *Stratified sampling* after selection of specific groups or geographical regions can be used to reduce variation. Figure 34 demonstrates the effect of sampling variation based on a study population of 10 animals where 50% of animals are of female sex. The estimates generated by 5 samples of 4 animals each vary between 25% and 75% female animals.

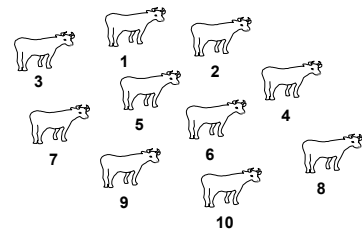
Population					
F					
M					
F					
M					
F					
M	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
F	F	M	F	F	F
M	M	M	M	M	F
F	M	F	F	M	F
M	M	M	M	F	M
50%	25%	25%	50%	50%	75%

Figure 34: Example of sampling variation

Probability sampling can be applied to the individual or the group as the sampling units. In the case of the first, the techniques *simple random*, *systematic* or *stratified sampling* are available. With *cluster* and *multistage* sampling sampling is applied at different levels of aggregation, and at least one of the sampling units has to be the group.

Simple random sampling

Strictly speaking, *simple random sampling* is the optimal method for selecting observations from a population. It is simple in theory, but can be difficult and not very efficient in practice. The assumption behind the procedure is that any of the possible samples from the study population has the same chance of being selected. This means that each individual has an equal probability of selection and an individual's selection does not depend on others being selected. Figure 35 presents an example of simple random sampling. A disadvantage of the technique is that it may result in large variation of the estimate thereby requiring larger sample sizes.



Take a simple random sample of 5 animals !

Sampling Frame based on Cow Numbers

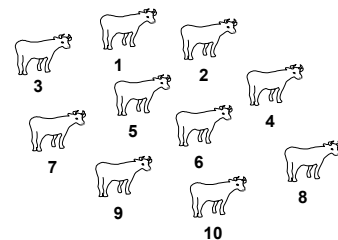
1	2	3	4	5	6	7	8	9	10
1	0	0	1	1	0	0	1	0	1

Random Number (1 selects)

Figure 35: Example of simple random sampling

Systematic random sampling

Systematic random sampling is a very practical method for obtaining representative samples from a population. It ensures that the sample is evenly distributed across the study population. Figure 36 illustrates the application of the technique. It can introduce selection bias, if the characteristic measured is related to the interval selected. This can be the case, if the sampling interval is subject to seasonal or behavioural effects. It is also mathematically more difficult to obtain valid variance estimates, but in practice simple random sampling estimators are being used.



Take a systematic random sample of 5 animals !

10 animals in sampling frame

$$5 / 10 = 2$$

take every second animal coming through gate

Figure 36: Example of systematic sampling

Stratified random sampling

Stratified random sampling is an effective method for reducing variance, if a known factor causes significant variation in the

outcome variable, but is not the target of the analysis. For example, in the case of milk production in a population of dairy cows of the Jersey and Holstein breeds, sampling variation of estimates will be substantial, largely due to genetic differences affecting milk volume between the two breeds. Stratification on breed will allow reducing the overall variation of the milk production estimate. Allocation of individuals to the different strata can be in equal numbers (same n per stratum) or proportional (same n/N per stratum). The latter is used if the researchers would like to ensure that the sample has a distribution of observations across strata which is representative of the target population. See Figure 37 for an example of applying stratified sampling. The technique will also allow easy access to information about the sub-populations represented by the strata. For stratified sampling to be effective at reducing variation, the elements within the strata should be homogeneous and variance between the strata should be large. As a disadvantage, one has to know the status of the sampling units with respect to the stratification factor and more complex methods are required to obtain variance estimates.

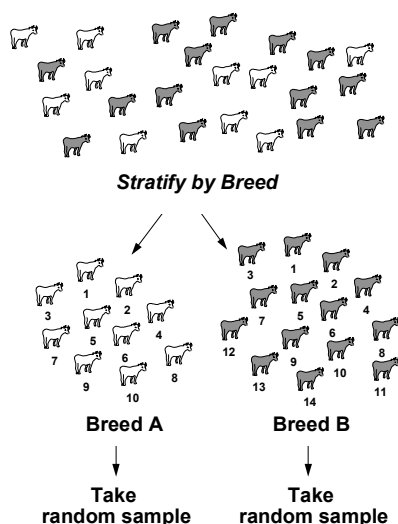
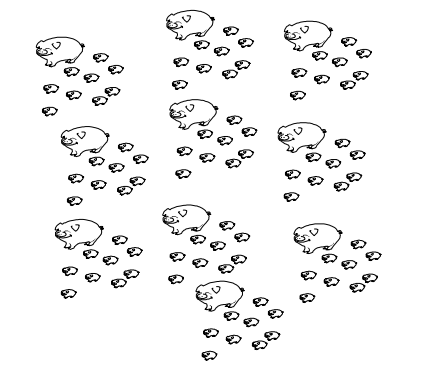


Figure 37: Example of stratified random sampling

Cluster sampling

Cluster sampling is one of the probability sampling techniques where sampling is applied at an aggregated level (=group) of

individual units. Typically, the individual still remains the *unit of interest* such as for example its disease status, but the *sampling unit* becomes a grouping of individual animals such as the herd or mob they belong to. All elements within each randomly selected group are then included in the sample. Therefore, this technique does only require a *sampling frame* for the groups, but not for the members within the groups. The groups or clusters can represent natural groupings such as litters or herds, or they can be based on artificial groupings such as geographic areas or administrative units. The random selection of the clusters as the sampling units can be performed using simple random, systematic or stratified random sampling. See Figure 38 for an example application of cluster sampling. With data collected on the basis of cluster sampling, the variance is largely influenced by the number of clusters, not the number of animals in the sample. The technique assumes that the elements within the different clusters are *heterogeneous* (unlike stratified sampling). It is often applied as part of multi-centre trials. Cluster sampling can lead to an increased sampling variance following the saying that “birds of a feather flock together”. In this situation, a larger sample size would be required to reduce variance to acceptable levels. Large variation between and small variation within clusters will result in biased parameter estimates.

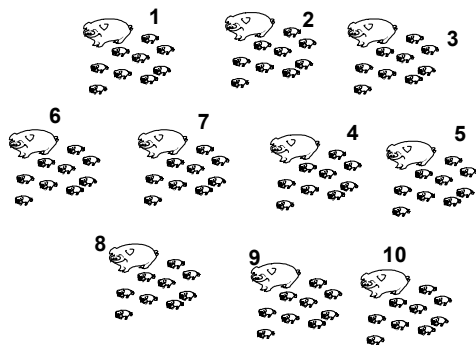


What is the prevalence of *E.coli* infection in total of 90 piglets?

Decide to take samples from about 30 piglets on economic grounds

9 piglets per sow

Random selection of 3 sows



Sows = Clusters

Sampling Frame based on Sow Numbers

1	2	3	4	5	6	7	8	9	10
0	0	0	1	0	0	0	1	0	1

Random Number (1 selects)

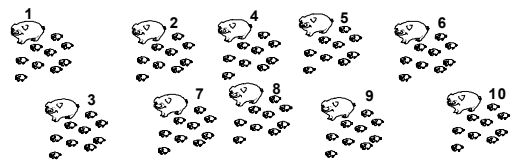
Select all piglets from litters of selected sows

Figure 38: Example of cluster sampling

Multistage sampling

Multistage sampling involves the use of random sampling at different hierarchical levels of aggregated units of interest. It is frequently applied as *two-stage sampling*, where herds are selected randomly as *primary sampling units* and within each of the selected herds animals are selected randomly as *secondary sampling units*. See Figure 39 for an example of the selection process. The optimal ratio between the number of primary and secondary sampling units can be determined on the basis of cost and/or variability between and within primary

sampling units. Multistage sampling is mainly used for practical reasons such as for example in situations where it is difficult to establish a sampling frame for the secondary sampling units. In this type of situation, after the primary sampling units such as for example the herds have been randomly selected and subsequently visited, a sampling frame for the secondary sampling units, say the dairy cows within each selected herd, can be established on the basis of each farmer's records. A disadvantage of this technique is that it can result in an increased sample size.



Sample size of 30 piglets
Random selection of 5 sows

Sampling Frame based on Sow Numbers

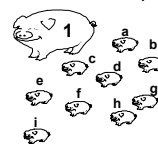
1	2	3	4	5	6	7	8	9	10
1	0	0	1	0	1	0	1	0	1

Random Number (1 selects)



Random selection of 6 piglets from litters of randomly selected sows

Example for Litter of Sow 1



Piglet	Random Selection
a	0
b	1
c	0
d	0
e	1
f	0
g	1
h	1
i	1
j	1

Figure 39: Example of multistage sampling

Multistage sampling is often used as part of epidemiological studies. In combination with stratified sampling the use of multistage sampling is recommended by the Office International des Epizooties as part of the official pathway to declaration of freedom from infection with the rinderpest virus (see Figure 40).

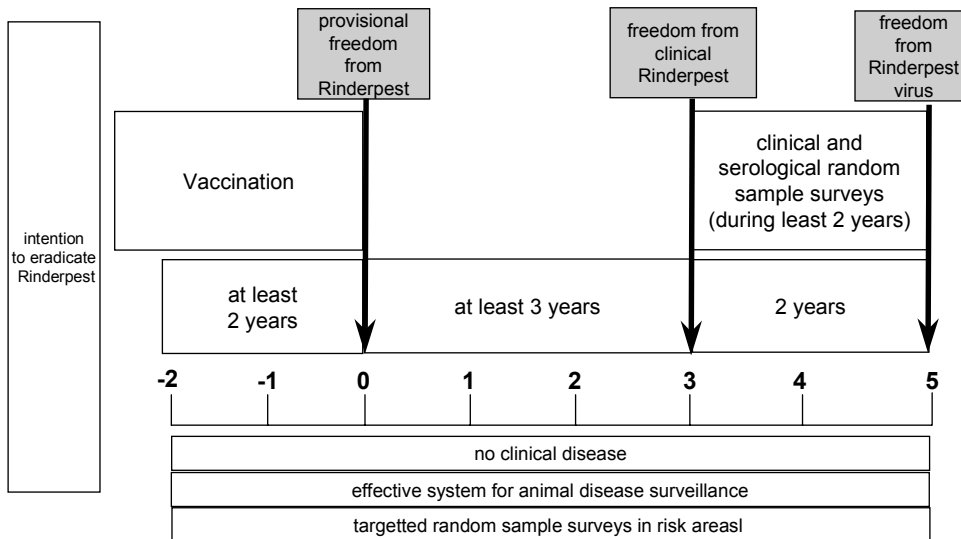


Figure 40: OIE pathway to declaration of freedom from rinderpest infection

Comparison of main sampling methods

The main sampling methods are compared in Figure 41 with respect to the population characteristics and the population categories for which a particular approach is most useful.

Population Characteristics	Population Type	Appropriate Sampling Technique
homogeneous	cattle on a farm, to be sampled to determine tuberculosis prevalence	simple random
definite strata, but homogeneous within strata	farm with 2 different dairy breeds, each with equal number; sampled to determine milk production	simple stratified
definite strata, each stratum has proportionate ratio of number of members of other strata	farm with 2 different dairy breeds, but different numbers; sampled to determine milk production	proportional stratified
groups with similar characteristics, but heterogeneous within group	veterinary laboratories in a country equipped according to a standard; wide variation between samples submitted; to be sampled to determine proportion of contaminated tissue samples	cluster
no sampling frame for units of interest	cattle in a region, to be sampled to determine tuberculosis prevalence	multistage

Figure 41: Comparison of main sampling methods

Sample Size Considerations

A decision as to the required sample size has to be based on different calculations depending on whether estimates for a *categorical* or a *continuous* variable are to be calculated. The factors to take into consideration include the *accuracy* required, the *sampling method* to be used, the *size of the smallest subgroup* and the *actual variability* of the variable of interest in the population.

As a very brief background on sampling theory, the *central limit theorem* is the basis of the variance estimates calculated for probability sampling data. The central limit theorem states that given large sample sizes the distribution of sample means or proportions tends to be normal. Using this theory, confidence intervals can be calculated with 90% of estimates ranging between +/- 1.65 standard errors and 95% between +/- 1.96 standard errors (see Figure 42). It should be remembered that confidence intervals are interpreted such that in the case of for example the 95% level 95 of the confidence limits calculated for 100 samples taken from the same population will include the true population value.

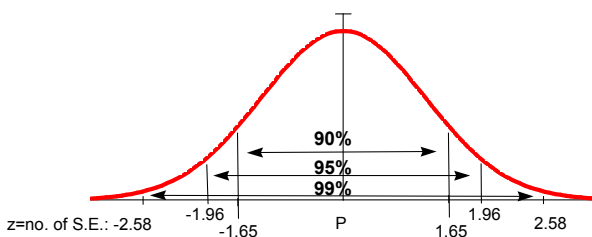


Figure 42: Central limit theorem and confidence limits

P	P (1-P)
0.5	0.25
0.4	0.24
0.3	0.21
0.2	0.16
0.1	0.09

Figure 43: Prevalence and sample size

Estimation of level of disease occurrence

If the objective is to estimate the disease frequency, the following information is

required. First, a guesstimate of the probable prevalence of reactors (P) has to be obtained. If it is not known, $P=0.5$ can be used as this gives the largest sample size given the same *absolute precision*. Then, a decision has to be made as to the desired confidence level (α) (for example 0.95). The z values to be used in the formula shown in Figure 44a are 1.65, 1.96 and 2.58 for 90%, 95% and 99% confidence levels respectively. Then, the level of precision d has to be decided. This parameter is the distance of the sample estimate in either direction from the true population proportion considered acceptable by the investigator. It can be expressed either as a number of percentage points (*absolute precision*) or a percentage of the expected value (*relative precision*). The *design effect* parameter should be applied to adjust for sampling design. In the case of stratified sampling variance typically has to be adjusted downwards and with cluster or multistage sampling variance has to be adjusted upwards using the design effect as a multiplier of estimated sample sizes. The formula for an infinite population assuming a 95% CI ($z=1.96$) and precision d with n being the required sample size is presented in Figure 44a. It contains the parameter $P(1-P)$ which as mentioned above is largest for a prevalence of 0.5 assuming the same value for d (see Figure 43). If the sample size n is greater than 10% of the total population size, use the formula described in Figure 44a to obtain the sample size n^* followed by applying the correction for finite populations presented in Figure 44b.

While it is useful to understand the principles behind these calculations, the required sample sizes can be obtained much more quickly from tables or specialised epidemiological computer software such as EpiInfo or EpiScope. Figure 44c presents a table of the sample sizes given a 95% confidence level for different prevalence and absolute precision levels. For example, to estimate the prevalence of disease in a large population to within +/- 5% at the 95% confidence level with an expected prevalence of 20%, it is necessary to examine a random sample of 246 animals.

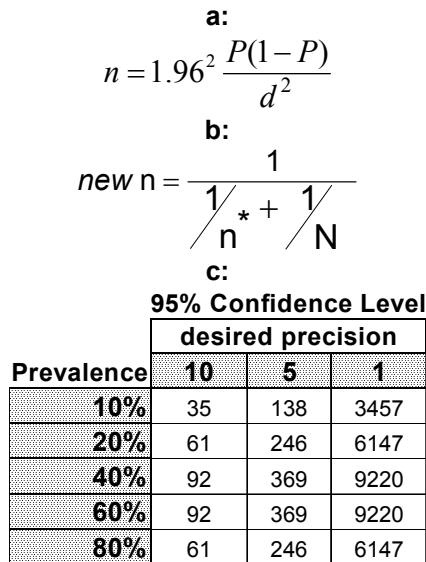
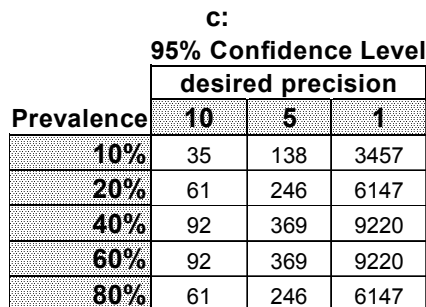


Figure 44: Estimation of level of disease occurrence

The following example will demonstrate the use of the formula. The investigator has been asked to determine the proportion of cull cows that will yield a positive culture for *M. paratuberculosis*. The acceptable absolute precision is +/- 5% and the expected prevalence when sampling at the slaughter house is assumed to be 10%. Applying the formula in



a indicates that about 138 cattle will have to be sampled to determine the proportion of *M. paratuberculosis* in cull cows (see Figure 45).

$$n = 1.96^2 * (0.10) * \frac{(1 - 0.10)}{(0.05)^2} = 138\ cattle$$

Figure 45: Example sample size calculation for disease prevalence estimation

Sampling to detect disease

During outbreak investigations, disease control/eradication programs or if testing the whole herd is too expensive, the objective is often to determine the presence or absence of disease. Figure 46a and b show the formulas

for finite as well as for infinite population sizes. To simplify the process as mentioned above computer programs can be used or a table such as the one shown in Figure 46c. The interpretation of the sample size obtained from the table is that if no animal in the sample tested gives a positive test result, you can assume with 95% level of confidence that the disease is not present in the population.

a: *Finite populations:*

$$n = \left[1 - (1 - \beta)^{\frac{1}{d}} \right] \left[\left(N - \frac{d}{2} \right) + \frac{1}{2} \right]$$

(β = confidence level (as proportion) -> probability of observing at least one diseased, if prevalence is d/N; N = population size, n=sample size; d = number of diseased)

b: *Infinite populations (> 1000):*

$$n = (\log(1 - \beta)) / \left[\log \left(1 - \frac{d}{N} \right) \right]$$

(n=sample size, β = level of confidence, d= number of diseased, N=population size)

c:

Population Size	Prevalence					
	0.1%	1%	2%	5%	10%	20%
10	10	10	10	10	10	8
50	50	50	48	35	22	12
100	100	96	78	45	25	13
500	500	225	129	56	28	14
1000	950	258	138	57	29	14
10000	2588	294	148	59	29	14
infinite	2995	299	149	59	29	14

If the expected prevalence is 5% and the population size is 500, a sample of 56 is required to be 95% certain of detecting at least one positive.

Figure 46: Formulas and table (95% confidence level) for sample size to detect presence of disease

Probability of not detecting disease

In the case of importation of animals, it may be necessary to quantify the probability of failure to detect any positives in a sample from an infinite population. The assumption for the formula is that population size is infinite and prevalence (prev) is given (see Figure 47).

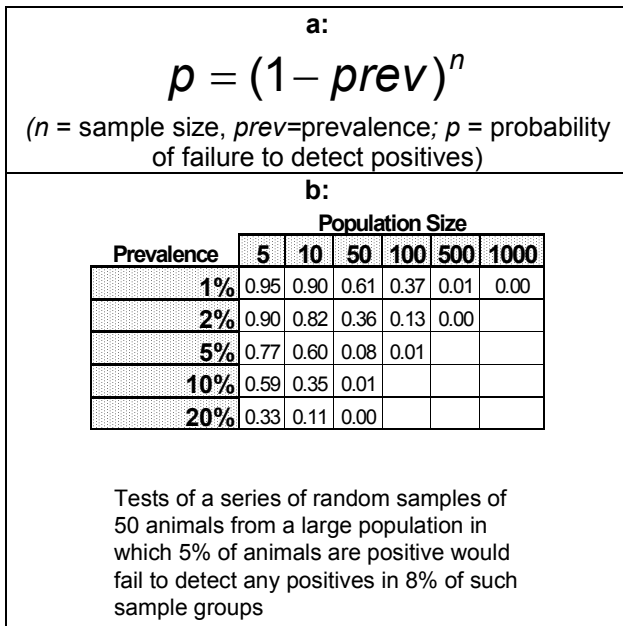


Figure 47: Formula and table for probability of not detecting disease

Simplified formula for disease detection sampling and sampling for disease missed

A simplified formula can be used for disease detection as well as the possible number of diseased animals missed during the sampling process. In the case of a 95% confidence level it is called the *Rule of Three*. To determine the required sample size for disease detection, 300 is divided by the expected proportion diseased. To determine the possible number of diseased animals in the population given that a specific sample size indicated that all animals in the sample were negative, 300 is divided by the number of animals tested to give the prevalence of diseased animals potentially missed. The number 300 can be replaced with 460 for 99% confidence and 690 for 99.99% confidence. As an example for disease detection using the simplified formula, a sample size is required so that the investigators can be 95% confident that no disease is present in the population. Assuming a herd size of 500 sheep and an expected prevalence of 25% for caseous lymphadenitis, 300 divided by 25 gives a minimum sample size of 12 animals. As an example for estimating the number of diseased animals missed using the simplified formula, a sample of 10 animals were all tested negative resulting in a prevalence of 30% or 150 diseased animals which potentially could still be present.

Sample size for estimation of continuous-type outcome variable

With continuous-type dependent variables an estimate of not only the expected value but also its variation is necessary. You can use the formula presented as Figure 48a for P= 0.05 ($z_{\alpha}=1.96$) and Power = 0.9 ($z_{\beta}=1.28$). As an example, let us assume that the investigator would like to estimate lambing-to-conception interval in sheep. The expectation is that about 2/3 of animals will be within 20 days on either side of the average (SD about 20 days). A precision of within 5 days of the true average has been requested. Using the formula presented in Figure 48b, the recommended sample size would be 64 animals.

a:

$$n = \left(\frac{(1.96 - 1.28)\sigma}{L} \right)^2$$

(S = estimated standard deviation of parameter of interest; L = how accurate estimate is supposed to be expressed in units of parameter of interest)

b:

$$n = \frac{4 * (20)^2}{5^2} = 64$$

Figure 48: Sample size formula and example calculation for continuous measures

Interpretation of Diagnostic Tests

Learning Objectives

At the completion of this topic, you will be able to:

- define and differentiate the concepts of sensitivity and specificity
- evaluate a test in terms of its sensitivity, specificity, and the overall misclassification
- calculate predictive value and explain how predictive value is determined by sensitivity, specificity and the prevalence of the condition being evaluated
- understand the concept of likelihood ratios
- interpret ROC curves
- understand the use and implications of multiple testing

Uncertainty and the Diagnostic Process

The duties of the veterinary profession include “to maintain and enhance the health, productivity and well-being of all animals” and “to prevent and relieve animal suffering” (from: A guide to professional conduct for veterinarians in New Zealand. New Zealand Veterinary Association, May 1991). In order to fulfil this duty, the veterinarian has to be able to diagnose disease or production problems as well as identify possible causes. *Diagnosis* is the basis for a decision, such as whether to treat (or implement a program) or to do nothing, to further evaluate, euthanase or to wait. The tools which the veterinarian uses to come to a diagnosis include factual knowledge, experience, intuition as well as diagnostic tests (see Figure 49). Correct use of these four mechanisms maximises the probability of a correct diagnosis. The uncertainty with regard to the effect of a treatment on a patient’s health made the ancient Greeks call medicine a *stochastic art*. Clearly, the main task of any veterinarian is to deal with the uncertainty of both, diagnosis and the outcome of treatment. It has been shown in studies of the medical profession that fear of personal inadequacy and failure in

reacting to this uncertainty is a common characteristic among physicians. This has become an even more important problem as our society becomes increasingly specialised and technological, relying on science rather than religion or magic to explain uncertainties.

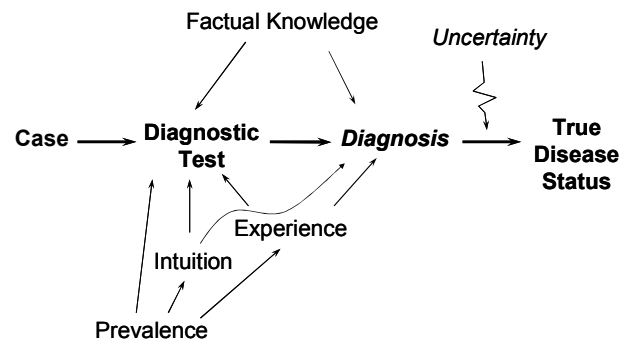


Figure 49: Factors influencing veterinary diagnoses

In this context, one should be aware of two major paradigms used to explain biological processes. The *mechanistic paradigm* assumes deterministic causation, and experiments are conducted to develop rules or laws according to which nature is thought to 'work'. Obviously, the client in a diagnostic situation does prefer this kind of interpretation, as it increases confidence in the diagnostic and the therapeutic process. The *probabilistic paradigm* on the other hand assumes probabilistic causation. Diagnostic and therapeutic procedures are seen as gambles and it is recognised that the decision-making process incorporates subjective judgement. The conclusion has to be though that given our incomplete understanding of biological systems and the presence of true biological variation, in veterinary diagnosis one must be content to end not in certainties, but rather statistical probabilities.

The outcome of the *diagnostic process* is a statement as to whether an animal is considered normal or not normal. This could relate to disease or infection status as well as to productive performance or quality of life from an animal welfare perspective.

Diagnostic Tests

The diagnostic test is a more or less objective method for reducing *diagnostic uncertainty*. As the consequential decisions are typically dichotomous (treat or do not), the outcome of the diagnostic process often is interpreted as a dichotomous variable as well, such as the animal having or not having the disease. The unit of measurement of the diagnostic device can be dichotomous, such as presence or absence of bacteria, which facilitates interpretation significantly. But if the diagnostic device measures on a continuous scale, such as serum antibody levels or somatic cell counts, a *cut-off value* has to be determined so that the result can be condensed into a dichotomous scale. Given a clinical measurement on a continuous scale, the problem with any cut-off point is that it is likely to result in overlap between healthy and diseased individuals with regard to test results (see Figure 50). The consequences of this situation are that uncertainty in addition to any other potential sources of measurement error (such as operator error) is being introduced. It is desirable to quantify this relationship between diagnostic test result and “true” disease status so that the clinician can take account of this uncertainty when interpreting test results.

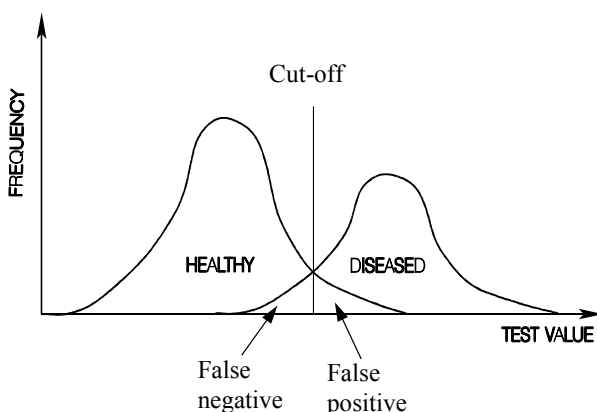
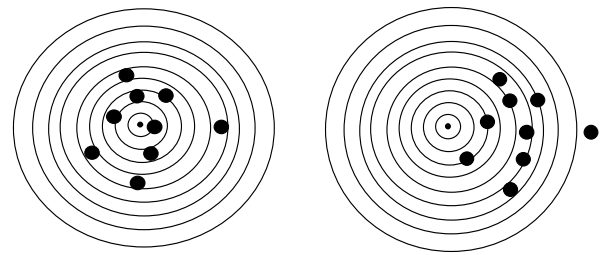


Figure 50: Test result measured on continuous scale

The performance of a diagnostic method can be described using the *accuracy* which refers to the closeness between test result and “true” clinical state, the *bias* which is a measure of the systematic deviation from “true” clinical

state, and the *precision* or *repeatability* representing the degree of fluctuation of a test series around a central measurement (see Figure 51).



imprecise

imprecise and biased

Figure 51: Precision and bias in diagnostic tests

Any evaluation of diagnostic tests needs a measure of the “true” condition of individuals to compare with which is usually called the *gold standard*. Most of the time it is impossible to define with 100% accuracy what the true diagnosis should be. There may also be disagreement amongst experts such as for example in the case of mastitis where the presence of the particular pathogen *or* the presence of an inflammatory response in the udder could be defined as the gold standard.

Evaluation and Comparison of Diagnostic Tests

To assess a diagnostic test or compare a number of different tests, it is necessary to apply the tests as well as the *gold standard* or *reference test* to a sample of animals from a population with a *typical disease spectrum*. But it may also be necessary to compare different methods without one of them being a gold standard. In this case one would assess the agreement between both diagnostic methods.

Comparison with gold standard

The clinician should be aware that the results of such evaluations could well differ between populations. The characteristics of the individual test relative to the gold standard are quantified through the sensitivity and specificity. *Sensitivity* defines the proportion of animals *with* the disease which test positive. In other words it is the ability to

correctly identify diseased animals and therefore gives an indication of how many false negative results can be expected. *Specificity* on the other hand is the proportion of animals *without* the disease which test negative. It represents the ability of the diagnostic tests to correctly identify non-diseased animals and gives an indication of how many false positive results can be expected. The two measures are inversely related and in the case of test results measured on a continuous scale they can be varied by changing the cut-off value. In doing so, an increase in sensitivity will often result in a decrease in specificity, and vice versa. The optimum cut-off level depends on the diagnostic strategy. If the primary objective is to find diseased animals meaning false negatives are to be minimised and a limited number of false positives is acceptable, a test with a high sensitivity and good specificity is required. If the objective is to make sure that every test positive is “truly” diseased (meaning no false positives, but limited amount of false negatives acceptable), the diagnostic test should have a high specificity and good sensitivity.

Agreement

Frequently in diagnostic test evaluation, no acceptable gold standard is available, and it may therefore become necessary to evaluate agreement between the tests, with one of the tests being a generally accepted diagnostic method. The *kappa test* is a statistical method for assessing the agreement between diagnostic methods measured on a dichotomous scale. It measures the proportion of agreement beyond that to be expected by chance. The statistic ranges from 0 to 1 with a kappa value of about 0.4 to 0.6 indicating moderate agreement. Higher kappa values are interpreted as good agreement. The kappa test can also be used to evaluate agreement between clinical diagnoses made by the same clinician on repeated occasions or between different clinicians.

The calculation of the Kappa statistic involves estimation of the observed proportion of agreement and the expected proportion assuming chance agreement as follows:

Observed proportion agreement: $OP = (a + d) / n$

Expected proportion agreement:

$$EP = \left[\frac{(a+b)}{n} \times \frac{(a+c)}{n} \right] + \left[\frac{(c+d)}{n} \times \frac{(b+d)}{n} \right]$$

$$\text{kappa} = (OP - EP) / (1 - EP)$$

The following example calculation for the kappa statistic is based on the comparison of longitudinal and cross-sectional examination of pig's heads for diagnosis of atrophic rhinitis (see Figure 52). The result indicates that there is moderate agreement between both types of examination. The longitudinal examination does seem to miss a substantial number of pigs ($n=16$) with atrophic rhinitis.

Longitudinal Examination	Cross-sectional Examination	
	Atrophy present	Atrophy absent
Atrophy present	8 (a)	1 (b)
Atrophy absent	16 (c)	223 (d)

$OP = (8 + 223) / 248 = 0.932$

$$EP = \left[\frac{(8+1)}{248} \times \frac{(8+16)}{248} \right] + \left[\frac{(16+223)}{248} \times \frac{(1+223)}{248} \right] = 0.874$$

$$\text{kappa} = (0.932 - 0.874) / (1 - 0.874) = 0.460$$

Figure 52: Example calculation for kappa statistic

Test Performance and Interpretation at the Individual Level

Predictive values are used when taking into account test characteristics during the diagnostic decision process. They quantify the probability that a test result for a particular animal correctly identifies the condition of interest. The predictive value of a *positive test* stands for the proportion of test positive animals which really have the disease. The predictive value of a *negative test* is the proportion of test negative which really do not have disease. Estimation of predictive values requires knowledge of *sensitivity*, *specificity* and the *prevalence* of the condition in the population. It is important to remember that predictive values are used for interpretation at the individual animal level and cannot be used to compare tests. The effect of prevalence on predictive values is considerable (see Figure 53). Given a 30% disease prevalence, 95% sensitivity and 90% specificity, the predictive value of a positive test would be 80% and for

a negative test 98%. If disease prevalence is only 3%, and test characteristics remain the same, the predictive value of a positive test will be 23% and for a negative test 99.8%.

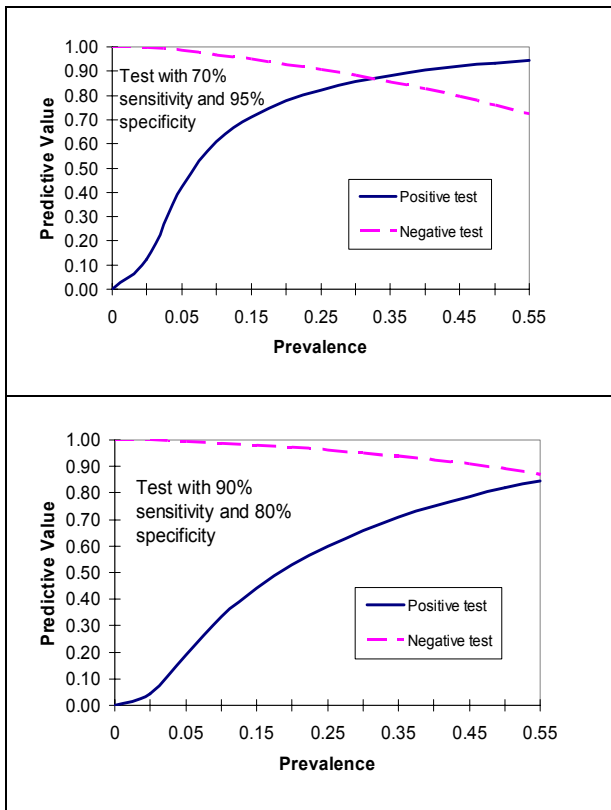


Figure 53: Relationship between prevalence and positive predictive value for tests with different sensitivity/specificity

Remember the following general rules about diagnostic tests:

- sensitivity and specificity are independent of prevalence (*note*: this is not necessarily correct)
- if prevalence increases, positive predictive value increases and negative predictive value decreases
- if prevalence decreases, positive predictive value decreases and negative predictive value increases
- the more sensitive a test, the better is the negative predictive value
- the more specific a test, the better is the positive predictive value.

Prevalence estimation with diagnostic Tests

Tests produce false negatives and false positives, therefore any diagnostic test can only produce an estimate of the apparent prevalence. The *apparent prevalence* is the proportion of all animals that give a positive

test result. It can be more than, less than or equal to the true prevalence. Estimates of the *true prevalence* can be obtained taking account of test sensitivity and specificity using the formula presented in Equation 1.

$$true\ prevalence = \frac{apparent\ prevalence + (specificity - 1)}{specificity + (sensitivity - 1)}$$

Equation 1: Formula for estimation of true prevalence

Calculations for evaluation of tests and test results

The different parameters which can be calculated for diagnostic tests and their results are summarised in Figure 54. Rather than memorising the calculations it is easier to work through them on the basis of the relationship between test results and true disease status using a 2-by-2 table layout. Even if no information about the actual numbers of animals in each cell is available, the table can be populated with proportions to calculate positive and negative predictive values, as long as prevalence, sensitivity and specificity are known.

	DISEASE	NO DISEASE	TOTAL
TEST POSITIVE	a	b	a + b
TEST NEGATIVE	c	d	c + d
TOTAL	a + c	b + d	N

$$SENSITIVITY = \frac{a}{a + c}$$

$$SPECIFICITY = \frac{d}{b + d}$$

$$POSITIVE\ PREDICTIVE\ VALUE = \frac{a}{a + b}$$

$$NEGATIVE\ PREDICTIVE\ VALUE = \frac{d}{c + d}$$

$$APPARENT\ PREVALENCE = \frac{a + b}{N}$$

$$TRUE\ PREVALENCE = \frac{a + c}{N}$$

Figure 54: Formulas for comparison and interpretation of diagnostic tests

Strategies for selection of an appropriate test

If the objective of diagnostic testing is to *rule out* disease, it means that a reliable negative result is required and therefore the test should generate few false negatives (=high *sensitivity*). In contrast, in order to *find evidence* (=rule in) of true disease and minimise false positive results, a reliable positive result is required with few false positives (=high *specificity*). The clinician interested in making *accurate* diagnoses should also recognise the importance of the *prior probability* of disease. If the *prior probability* is extremely small, a positive test result is not very meaningful and must be followed up by a highly *specific* test. The following rules of thumb can be used for successful interpretation of diagnostic test results. After the clinical examination, the *pre-test probability* should be estimated which may have to be revised in the light of new information. Then, before testing but after the clinical examination, the clinician should decide on an *action plan* for positive as well as negative test results. If both action plans are the same there should be no need for using the test. If the objective is to confirm likely diagnosis (“rule-in”), a test with at least 95% specificity and 75% sensitivity is required. If the sample tests positive, the positive predictive value will be high, which means that the animal is likely to have the disease. In the case of a negative test result, further diagnostic work-up is required. To confirm that an animal is free from disease (“rule-out”), the diagnostic test should have at least 95% sensitivity and 75% specificity. If the test result is negative, the negative predictive value will be high, meaning that the animal is likely to not have the disease. In the case of a positive test result, additional more specific tests are required.

Example of interpretation of diagnostic test results

(adapted from Baldock, C. 1988: Clinical epidemiology – Interpretation of test results in heartworm diagnosis. In *Heartworm Symposium*. Proceedings 107, Post Graduate Committee in Veterinary Science, University of Sydney, 29-33)

The diagnostic question is whether a particular dog does have an infection with adult *Dirofilaria immitis*? As a general strategy, the dog should first be evaluated clinically for presence of clinical symptoms of heartworm disease. This is followed by a microfilarial examination. If *microfilariae* are present, no further test is required. If no *microfilariae* are detectable and the clinical history is suggestive of heartworm disease, the diagnostic procedure should be followed by a serological test. The following 4 hypothetical case scenarios will demonstrate the decision process evolving from particular findings at different stages of the examination.

As **case 1**, a 5 year old untreated dog has been presented at a Brisbane clinic. The prevalence in the local dog population is 50% which serves as the first estimate of pre-test probability. During the clinical examination, the animal is found to have a history and clinical signs consistent with heartworm disease. As a consequence of this new information, the pre-test probability is revised to 80%. During the examination of the blood smear *microfilariae* are found. A differentiation from non-pathogenic filarids *D. repens* and *Dipetalonema reconditum* is performed on the basis of morphology and staining characteristics with acid phosphatase. If they are found to be *D. immitis*, a diagnosis of dirofilariasis is made and no further testing is required.

Case 2 is another 5 year old untreated dog presented at a Brisbane clinic with the same prevalence of 50% in the dog population (=first estimate of pre-test probability). History and clinical signs appear to be consistent with heartworm disease which results in a revised pre-test probability of 80%. On examination of the blood smear no *microfilariae* are found, and consequentially

the pre-test probability is revised to 60%. Because of the pre-test probability of 60% the dog is likely to have the disease, the diagnostic test is used to “rule-in” the disease. The serological test result is positive, which increases the probability of heartworm disease to above 95%.

Case 3 again is a 5 year old untreated dog presented at a Brisbane clinic with 50% prevalence in the local dog population (=first estimate of pre-test probability). Because history and clinical signs are consistent with heartworm disease, the clinician revises the pre-test probability to 80%. Examination of the blood smear does not show any evidence of *microfilariae*. The pre-test probability is therefore revised to 60%. The serological test is applied with the objective in mind to “rule-in” disease. The result from the serology is negative, which means that the probability of the dog not having heartworm disease becomes about 80%. To confirm that the dog does not have the disease, a “rule-out” test with high sensitivity should be used to confirm the result.

As **case 4**, a 5 year old untreated dog is presented at a Brisbane clinic because the owner would like to take the dog to New Zealand, and have the clinician provide a certificate that the dog is free from *D.immitis* infection. As above, the local infection prevalence is about 50% (= first estimate of pre-test probability). No clinical work-up and no blood smear is used as part of the diagnostic process. A serological test is used with the objective to “rule-out” *Dirofilaria immitis* infection. The clinician has the choice between different test kits. The operating characteristics of the serological tests are presented in Table 4. A comparison of the tests can be based on estimating the predictive values for positive and negative test results.

Table 4: Characteristics of serological test kits for *D.immitis* infection

Product	Sensitivity (in brackets number of diseased dogs in trial)	Specificity (in brackets number of non- diseased dogs in trial)
Dirokit Latex (2 trials)	85.3% (34) + 82.9% (76)	95.7% (23) + 100% (30)
Diromail (ELISA)	92.2% (103)	97.4% (78)
Dirocheck (ELISA)	92% (82) + 73%	99% (91) + 94%
Filarocheck (ELISA)	97.3% (149)	98.2% (165)
CITE (ELISA)	85% (266)	100% (176)

$$PPV = \frac{\text{a) } sensitivity * prevalence}{(sensitivity * prevalence) + (1 - specificity) * (1 - prevalence)}$$

$$NPV = \frac{\text{b) } specificity * (1 - prevalence)}{specificity * (1 - prevalence) + (1 - sensitivity) * prevalence}$$

Equations 2: Formulas for calculating positive and negative predictive values

The appropriate formulas are shown as Equations 2. The resulting predictive values are presented in Figure 55. Particularly from the perspective of the country allowing the importation of this dog, in this situation the predictive values of a negative test result are important. The results suggest that the Dirocheck kit assuming that the sensitivity / specificity values from the second trial are correct, performs rather poorly with a chance of 78% that a negative test result truly is negative. The Filarocheck kit would appear to provide the best performance, and the sensitivity/specificity data is based on reasonable sample sizes of dog populations. The possibility remains though that the dog population used in the evaluation of this particular test while large in size may in fact not be representative of the typical characteristics of the dog population in Brisbane.

PRIOR PROBABILITY	0.5			
TEST	SENSITIVITY	SPECIFICITY	PPV	NPV
DIROKIT	0.85	0.96	0.95	0.87
	0.83	1.00	1.00	0.85
DIROMAIL	0.92	0.97	0.97	0.93
DIROCHECK	0.92	0.99	0.99	0.93
	0.73	0.94	0.92	0.78
FILAROCHECK	0.97	0.98	0.98	0.97
CITE	0.85	1.00	1.00	0.87

Figure 55: Calculations for *D.immitis* serological diagnosis example (NPV = negative predictive value)

Methods for choosing Normal/Abnormal Criteria

The criteria for deriving cut-off values from diagnostic devices measuring quantities on a continuous scale such as optical densities of ELISA tests can be based on a range of different methods. The most popular technique is called the Gaussian distribution method, but in addition there are the percentile, therapeutic, risk factor, diagnostic or predictive value and the culturally desirable methods.

The *Gaussian distribution method* is used to derive a cut-off value on the basis of test results from a disease-free population. A histogram is drawn to confirm the Gaussian shape of the data, and the mean and standard deviation of the test results are computed. The upper and lower limits of the test results are defined using 2 standard deviations. That means 95% of values in the disease-free population will have values within this interval. The other values would be classified as abnormal. The advantage of the technique is that it is simple. But there are many disadvantages. Firstly, the distribution of values is likely to be skewed or bimodal. In addition, it is assumed that prevalence is fixed whereas in reality it will often vary between populations. There is also no biological basis for defining disease on the basis of such a cut-off. True normal ranges will differ between populations, and the approach does not recognise that changes over time in normal values can be pathologic.

With the *percentile method*, test values are obtained for a large number of disease-free animals, and the lower 95% are classified as normal, the upper 5% as abnormal. It is

possible to use the lower 2.5% and higher 2.5% instead. The percentile method is as simple as the Gaussian, but has the additional advantage that it is also applicable to non-normal distributions. Its disadvantages are otherwise the same as for the Gaussian method.

In the case of the *therapeutic method*, the cut-off value is decided on the basis of the level, at which treatment is recommended. New results from research will allow adjustment of the value. Its advantage is that only animals which are to be treated will be classified as diseased. A major disadvantage is its dependence on knowledge about therapeutic methods which has to be up-to-date.

The *risk factor method* uses the presence of known causally or statistically related risk factors to determine disease status. It has an advantage if risk factors are easy to measure and it facilitates prevention. But as a disadvantage, risk of disease may increase steadily (dose-response) which means that it becomes difficult to determine a cut-off. Therefore, the positive predictive value may be low.

Finally, there is the *diagnostic or predictive value method* which is considered the most clinically sound approach. With this technique, the cut-off is selected so that it produces a desired sensitivity and specificity. This can be done on the basis of the information contained in a *receiver operating characteristic* (ROC) curve. The choice of a suitable cut-off can be influenced by whether false-positives or false-negatives are considered less desirable by the clinician. The advantages of the predictive value method include that it can be applied to any value distribution. It uses realistic, clinical data for the development of the cut-off values. At the same time, it uses information about the diseased as well as the non-diseased population, and most importantly it can be adjusted to suit particular diagnostic objectives. The disadvantage of the method is that it requires monitoring of prevalence, positive and negative predictive values, and it could be seen as a disadvantage that the

clinician has to choose between a range of cut-offs.

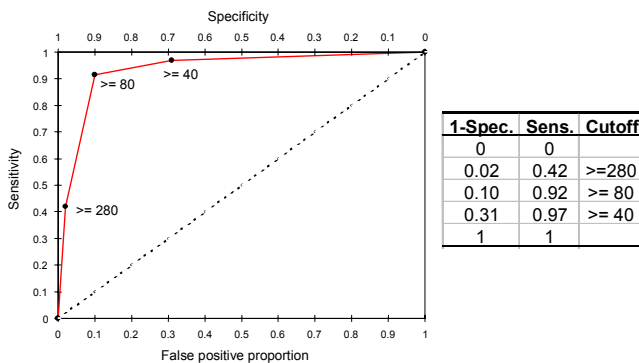


Figure 56: Example of receiver operating characteristic curve

The *receiver operating characteristic (ROC) curve* mentioned above in the context of the predictive value method consists of a plot of sensitivity and specificity pairs for different cut-off values (see Figure 56). The paired sensitivity/specificity values require that a dataset is available with test values and true disease status for a reasonable number of animals. The procedure is started by interpreting the individual animal test values using a particular cut-off value, and calculating the sensitivity / specificity pair for that cut-off using a 2-by-2 table. Using the example in Figure 5, interpreting the underlying individual animal data a cut-off value of ≥ 280 resulted in a sensitivity of 42% and a specificity of 98%. This process is then repeated for each cut-off value of interest, e.g. ≥ 80 and ≥ 40 in the example. Once this is done, one should have a whole series of sensitivity and specificity pairs (one for each cut-off), as presented in the data table in Figure 5. These pairs are then plotted as points on an X-Y chart with 1-specificity (=probability of false positive) on the X-axis and sensitivity on the Y-axis. The points are connected in sequential order of change in cut-off value to give the ROC curve. This analysis can now be done by computer software which will automatically assess the range of different cut-offs and plot the curve. The perfect diagnostic test should have 100% sensitivity and 0% false positives, and therefore should reach the upper left corner of

the graph. A diagonal ROC curve (from lower left to upper right corner) indicates a diagnostic test which does not produce any useful differentiation between disease and non-diseased. The area under the ROC curve can be used to quantify overall test accuracy. The larger this area, the better the test. Different diagnostic tests can be compared by plotting their ROC curves. The test whose ROC curve comes closest to the upper left corner is the best test. The ROC Curve can be used to adjust cut-off values according to different diagnostic strategies as follows. If false-negatives and false-positives are equally undesirable, a cut-off on the ROC curve should be selected which is closest to the upper left corner of the X-Y chart. Given that false-positives are more undesirable, a cut-off further to the left on the ROC curve should be used. In case that false-negatives are more undesirable, the cut-off should be set to a value towards the right on the ROC curve

Likelihood Ratio

The purpose of a diagnostic test is to influence the clinician's opinion about whether the animal has the disease or not. This means that it is used to modify the opinion (this better be a probability!) which the clinician had before obtaining the test result, i.e. it is a modifier of the *pre-test probability*. For this purpose, presenting the test result as positive or negative will not be that useful, because the result could still be a false-positive or negative. A quantity is therefore needed which gives the clinician an idea how *likely* it was that the test result could have been produced by a diseased compared with a non-diseased animal. This can be done with the *likelihood ratio*. They can be calculated for negative as well as positive test results. The likelihood ratio for a *positive test* is estimated on the basis of dividing the probability of a particular test result in the presence of disease (=sensitivity) by the probability of the test result in the absence of disease (=1-specificity). The result is interpreted as how *likely* it is to find a positive test result in diseased compared with non-diseased individuals. The likelihood ratio of a *negative test result* is calculated as the

quotient between (1-sensitivity) and specificity (see Figure 57).). It is used less frequently than the one for a positive test. This result is then interpreted as how *likely* it is to find a negative test result in diseased compared with non-diseased animals.

Likelihood ratios (LR) can be calculated using single cut-off values, so that one obtains only one pair of likelihood ratios, one for a positive (LR+) and another for a negative test result (LR-). More powerful information can be extracted from the diagnostic test by using *multilevel likelihood ratios*. In this case every test value, or more often several ranges of test values, will have a specific LR+ and LR-. The advantage of the *multilevel LR* method is that it allows the clinician to take account of the degree of abnormality, rather than just use crude categories such as presence or absence of disease. It will be possible to place more emphasis on extremely high or low test values than on borderline ones, when estimating the probability of disease presence (i.e. *post-test probability*). So for example a LR values around 1 will not allow much differentiation between diseased and non-diseased animals. These values are often generated for test values around the cut-off point which would have been used with the traditional positive/negative interpretation of a diagnostic test. On the other hand, an LR of 10 will provide much stronger indication of disease presence, and would typically be obtained for test values further away from the aforementioned cut-off value. The *multilevel LR* values for each range of test values are calculated as the ratio between the proportion of diseased animals and the proportion of non-diseased animals within that particular range of test values.

In essence, with both types of likelihood ratio, the interpreted output of the diagnostic device for a particular sample will become a likelihood ratio value rather than a test-positive/negative assessment.

Likelihood ratios (LR) do not depend on prevalence, and they provide a quantitative measure of the diagnostic information contained in a particular test result. If used in

combination with the initial expectation of the probability that an animal has a certain condition (= *pre-test probability*), a revised estimate of the overall probability of the condition (= *post-test probability*) can be calculated. A complete set of *multilevel likelihood ratios* does, in fact, contain the same information as a ROC curve.

In order to perform the revision of the *pre-test probability*, it has to be converted into a *pre-test odds* as described in Figure 57a. The result is then multiplied with the likelihood ratio for a particular test value to produce an estimate of the *post-test odds*. This in turn then has to be converted back into a probability as shown in Figure 58b.

$$LR (+) = \frac{\text{sensitivity}}{1 - \text{specificity}}$$

$$LR(-) = \frac{1 - \text{sensitivity}}{\text{specificity}}$$

Figure 57: Formulas for likelihood ratio calculations

$$a) \quad \text{odds} = \frac{\text{probability of event}}{1 - \text{probability of event}}$$

$$b) \quad \text{probability} = \frac{\text{odds}}{1 + \text{odds}}$$

Figure 58: Calculation of pre-test odds and post-test probability

As an example, somatic cell counts (SCC) are used as a screening test for sub-clinical mastitis. A particular dairy client is known to have a mastitis prevalence of about 5% in the cow herd. One cow shows an SCC of 320. The probability that the cow does indeed have mastitis can be estimated based on a fixed SCC threshold or using the likelihood ratio for a positive test. Using the first method, assuming a threshold of SCC=200 and a sensitivity of 80% and a specificity of 80% a positive predictive value of 17% can be calculated (see Figure 59a). Using the likelihood ratios, the post-test probability becomes 43.5% which gives the test result a

much stronger likelihood of representing an actual case of mastitis (see Figure 59b).

a: Calculation using fixed threshold:

	MASTITIS	HEALTHY	TOTAL
ELEV.SCC	40	190	230
LOW SCC	10	760	770
TOTAL	50	950	1000

Positive predictive value

$$= \frac{40}{230} = 0.17 \approx 17\%$$

b: Calculation using likelihood ratio:

Table of likelihood ratio

SCC	< 100	100-200	200-300	300-400	>400
LR (+)	0.14	0.37	2.5	14.5	40.8

pre-test odds = 0.05 / 0.95 = 0.053
 post-test odds = 0.053 x 14.5 = 0.769

Post-test probability

$$= \frac{0.769}{1 + 0.769} = 0.435 \approx 43.5\%$$

Figure 59: Example calculations for somatic cell count result interpretation

The calculation of post-test probabilities can be greatly facilitated by using a nomogram as shown in Figure 60. In the example present in this figure, a test result was produced which was presented as a likelihood ratio value for a positive test result of 10. The pre-test or prior probability as a result of knowledge of the population prevalence or clinical examination is assumed to be 20%. The post-test probability was obtained by drawing a straight line through from the pre-test probability via the LR value through to the resulting post-test probability of 75%.

**Example:
 Prior probability = 20%
 LR = 10**

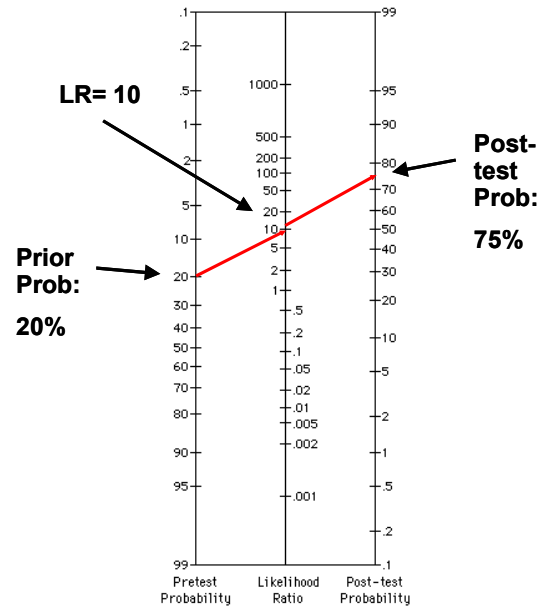


Figure 60: Nomogram for post-test probability calculation using likelihood ratios of a positive test result

The relationship between pre-test probability and various LR values is summarised in Table 5. It clearly demonstrates that with a low pre-test probability such as 5%, it is difficult even with a fairly high LR of 10 to be confident enough to diagnose an animal as diseased. In fact, given a likelihood ratio of 3, one would need at least a 50% pre-test probability to be more than 50% confident that an animal has the disease.

Table 5: Post-test probabilities for different pre-test and likelihood ratio value combinations

		Pre-test probability					
		5%	10%	20%	30%	50%	70%
Likelihood ratio	10	34%	53%	71%	81%	91%	96%
	3	14%	25%	43%	56%	75%	88%
	1	5%	10%	20%	30%	50%	70%
	0.3	1.5%	3.2%	7%	11%	23%	41%
	0.1	0.5%	1%	2.5%	4%	9%	19%

Combining Tests

Different diagnostic methods are frequently used in combination to allow improved diagnosis through strategic decisions about

interpretation of results. These approaches include use of different tests for the same disease problem on a single animal, of different tests each identifying different disease problems in a single animal or of the same test for a specific disease problem applied to several animals or the same animal over time.

Different tests for the same disease problem in a single animal

Different diagnostic methods are frequently used in combination to increase the clinician's confidence in a specific diagnosis in a particular animal. But if this is done purely in a qualitative manner, it is very difficult to fully take advantage of the combined information. Different approaches can be used to integrate information provided by the individual test results, in particular *parallel* and *series* interpretation are important. In this context it is essential to recognise that these methods relate to usage of different tests at the same time in a single animal. The tests should be different, in that they measure different biological parameters, e.g. one may be an ELISA test for tuberculosis and the other the tuberculin skin test.

Parallel interpretation

With parallel test interpretation, an animal is considered to have the disease, if one or more tests are positive. This means the animal is being asked to 'prove' that it is healthy. The technique is recommended if rapid assessment is required or for routine examinations, because the animal is considered positive after first test since in that case the second test result does not matter anymore. If the first test is negative, the second test is still necessary. Parallel test interpretation will increase sensitivity and the predictive value of a negative test result, therefore disease is less likely to be missed. But on the other hand it does reduce specificity and the predictive value of a positive test, hence false-positive diagnoses will be more likely. As a consequence, if you conduct enough tests an apparent abnormality can be found in virtually every animal even if it is completely 'normal', since individual tests are only rarely 100% specific.

Series interpretation

With serial test interpretation, the animal is considered to have the disease if all tests are positive. In other words, the animal is being asked to 'prove' that it has the condition. Series testing can be used if no rapid assessment is necessary because for an animal to be classified as positive all test results have to be obtained. If some of the tests are expensive or risky, testing can be stopped as soon as one test is negative. This method maximises specificity and positive predictive value, which means that more confidence can be attributed to positive results. It reduces *sensitivity* and negative predictive value, and therefore it becomes more likely that diseased animals are being missed. Likelihood ratios can be applied to the interpretation of serial testing by using the post-test odds resulting from a particular test as the pre-test odds for the next test in the series.

The effects of parallel and series testing are compared in Table 6. In this example, a disease prevalence of 20% is assumed. Test A has a moderate sensitivity and a poor specificity resulting in a very poor positive predictive value. Test B has good sensitivity and specificity producing a better, but still less than desirable, positive predictive value. A series interpretation of the test results substantially improves the predictive value of positive test results, to 82% where they become useful from a diagnostic perspective. A parallel interpretation of the two test results improves the predictive value of negative tests from 92% for test A and 97% for test B to a level of 99%.

Table 6: Example of parallel and series interpretation of a diagnostic test

Test	Sensitivity (%)	Specificity (%)	Positive predictive value (%)*	Negative predictive value (%)
<i>A</i>	80	60	33	92
<i>B</i>	90	90	69	97
<i>A and B (parallel)</i>	98	54	35	99
<i>A and B (series)</i>	72	96	82	93

*assuming 20% prevalence

Screening and confirmatory testing

With a strategy of screening and confirmatory testing, as it is often used in a disease control scheme, the screening test is applied to every animal in the population to search for test-positives. This test should be easy to apply at a low cost. It has to be a highly sensitive test so that it misses only a small number of diseased or infected animals. Its specificity should also be reasonable, so that the number of false positives subjected to the confirmatory test remains economically justifiable. Negative reactions to the screening test are considered definitive negatives, and are not submitted to any further tests. Any positive screening test result is subjected to a confirmatory test. This test can require more technical expertise and more sophisticated equipment, and may be more expensive, because it is only applied to a reduced number of samples. But it has to be highly specific, and any positive reaction to the confirmatory test is considered a definitive positive.

Comparison of combined test interpretation strategies

The different strategies for combining multiple test results are compared in Table 7.

Table 7: Comparison of methods of combining test results

Considerations	Test strategy		
	Parallel	Series	Screening/ confirmation
Effect of strategy	increase SE	increase SP	increase SE and SP
Greatest predictive value	negative test result	positive test result	both
Application and setting	rapid assessment of individual patients; emergencies	diagnosis when time is not crucial; test and removal programmes	Cost – effectiveness; test and removal programmes
Comments	useful when there is an important penalty for missing disease	useful when there is an important penalty for false positives	useful when large numbers of animals are to be tested

Using the same test multiple times

The same test can be used in several related animals at the same time (*aggregate testing*), of the same herds at different times (*negative-herd retesting*) or in the same animal at different times (*sequential testing*). The first and second approach is particularly important with disease control programmes, whereas the third, for example, may be used as part of research programmes.

Aggregate testing

Most animal disease control programmes rely on testing of aggregate populations, such as herds. The success of such programmes can be influenced greatly by the performance of individual tests, but also by their interpretation at the herd level. It is important to recognise that with animal disease control programmes there are two units to be diagnosed, the herd and the individual animal within the herd. This approach takes advantage of the epidemiology of an infectious organism, where in most cases spread is more likely within than between herds. In practical terms, this means once an infected animal has been detected within a herd, specific measures are being taken which are aimed at restricting the possibility of spread to other herds with susceptible animals. These may involve that none of the animals from this herd can be moved from the property, or that all of them are being culled, even though only one animal reacted positive to the test.

Tests with moderate sensitivity or specificity at the individual animal can be used due to the aggregate-level interpretation of the test (herds). The interpretation of the test results from individual animals within the herds is usually done in *parallel*, in order to achieve a increase sensitivity at the aggregate level while a less than optimal specificity is often considered acceptable. Detection of at least one positive animal will therefore result in the herd being classified as positive. There are some diseases where the presence of positive animals up to a certain number may still be classified as negative, since test specificity below 100% will always produce false-positives.

It has to be taken into consideration that herd size will influence the probability of identifying infected herds, because the more animals are being tested, the more likely it becomes to detect true positives as well as false positives. When disease control programmes commence, prevalence of infected herds can be above 20%. At this stage, the apparent prevalence will be higher than the true prevalence, as a consequence of less than 100% specificity of the diagnostic method. The lower the prevalence becomes, the larger the gap will be between apparent and true prevalence. Therefore, the predictive value of positive test results (at aggregate or individual animal level) will decrease and the proportion of false positives will increase. During the early phase of a control program, *sensitivity* is more important than specificity in order to ensure that all infected animals within infected herds are being detected. During this phase, there will also be a larger number of herds with high prevalence compared with the later phases of the control program. As the prevalence decreases, *specificity* becomes as important as sensitivity, and it may become necessary that a second test is carried out using series interpretation to further increase specificity (see screening/confirmatory testing above). It should be noted though that *sensitivity* will still be important during this phase, as the control program cannot afford missing too many infected animals.

Negative-herd re-testing

Within the context of animal disease control programmes, negative-herd re-testing is a typical testing strategy. This means that only animals that were negative to an initial test undergo a further test, because any positive animals would have been culled. Interpretation of results is usually at the herd level. This increases aggregate-level (= herd level) sensitivity because, if there are diseased animals in the herd, even a test with a moderate sensitivity will eventually detect at least one of them. The testing strategy increases the chance of finding infection missed on previous testing rounds. In principle, the herd is asked to 'prove' that it is

free from the condition. With decreasing prevalence in the population, specificity becomes more important. For example in a disease-free population, a test with 80% specificity keeps producing a 20% prevalence.

Sequential testing

Sequential testing is used as part of specific studies where one has the opportunity to repeatedly test the same animal over time to detect seroconversion. This technique is quite powerful, as it does not rely on a single result for interpretation, but rather on a significant change in test value which may well remain below a cut-off which would otherwise be classified as non-diseased on the basis of single samples. This type of testing could be used for example during experimental infection studies.

Using different tests for different disease problems in the same animal

Batteries of multiple tests have become quite common in small animal practice, where a blood sample from a single animal is sent to a laboratory for assessment of different blood metabolite levels. The objective is to identify normal and abnormal parameters. Testing for allergic reactions is an example of this. The technique becomes useful if a set of different parameters is of diagnostic value for establishing a pattern that is considered suggestive of a particular disease. The approach becomes questionable, however, if it is part of a 'fishing expedition' for a diagnosis. The clinician has to keep in mind that a cut-off for a single test is typically set such that it includes 95% of the normal population, which means it will produce 5% false positives. As an example, with a battery of 12 diagnostic tests measuring different blood parameters, each of them will have a 0.95 probability of diagnosing a 'normal' animal correctly as negative. But it also means that the overall chance of a correct negative diagnosis on all tests is $(0.95)^{12}$ amounts to a probability of 0.54. Therefore there is a 46% chance that a 'normal' animal has at least one false-positive result among these 12 tests.

Decision Analysis

Decisions always have consequences and these can be very important. Decision problems become difficult if they are complex, require multiple successive decisions and each decision possibly has more than one outcome. In the presence of uncertainty about the outcome of a decision, the decision maker is in fact forced to gamble. In the case of veterinarians working in general practice, clinical decisions are continually made under conditions of uncertainty. The latter being caused by *errors* in clinical and lab data, *ambiguity* in clinical data and *variations* in interpretation, *uncertainty* about *relationships* between clinical information and presence of disease, *uncertainty* about *effects* and costs of treatment and *uncertainty* about *efficacy* of control procedures and medication. Other areas where decision problems are particularly complex include the planning of disease control policies. In this situation, uncertainty is introduced for example through incomplete knowledge of the epidemiology of diseases or stochastic effects influencing disease spread. It is also not possible to predict the behaviour of individuals who are involved in handling and managing the disease vector or host.

Decision analysis is applicable to the decision problems described above. The process of *decision analysis* involves identifying all possible choices, all possible outcomes and structuring the components of the decision process in a logical and temporal sequence. *Decision tree analysis* uses a tree structure to present the different decision options and possible outcomes. The tree develops sequentially from *base* to *terminal* ends based on the components: *nodes*, *branches* and *outcomes*. There are three types of *nodes*: *decision* (choice) nodes, *chance* (probability) nodes and *terminal* nodes. The *branches* indicate the different *choices* if they are extending from a *decision* node and the different *outcomes* if they are extending from a *chance* node. Each of the *branches* emanating from a *chance* node has associated *probabilities* and each of the terminal ends has associated *utilities* or *values*. In the case

of decision trees a solution is typically obtained through choosing the alternative with the highest *expected* monetary value. This is done through *folding back the tree*. Starting from the terminal nodes and moving back to the root of the tree, expected values are calculated at each chance node as the weighted average of possible outcomes where the weights are the chances of particular outcome occurrences. At each decision node the branch with the highest expected value is chosen as the preferred alternative.

Decision tree analysis has the advantage that it encourages to break down complex problems into simpler components such as choices, probabilistic events and alternative outcomes. It encourages to weigh risks and benefits, logical sequencing of components and requires explicit estimates of probabilities. Concern about utilities is encouraged through the need of placing values on them. Critical determinants of the decision problem are identified and areas of insufficient knowledge are indicated.

As an example of a veterinary decision problem, a client has to decide whether to treat a cow valued at \$1000 diagnosed with traumatic reticulitis conservatively using a \$15 magnet or to spend \$150 on surgery. The decision tree is presented in Figure 61. The assumptions are being made that the probability of recovery is 0.9 for surgical and 0.8 for magnet treatment. The salvage value of the cow amounts to about \$400. The expected monetary values for the two treatments are calculated as follows:

- expected value for surgery

$$EV_{surgery} = 0.9 \times (\$1000 - \$150) + 0.1 \times (\$400 - \$150) = \$790$$
- expected value for magnet treatment

$$EV_{magnet} = 0.8 \times (\$1000 - \$15) + 0.2 \times (\$400 - \$15) = \$865$$

The interpretation of these results is that *in the long run* the magnet treatment is more profitable assuming that the values and probabilities are chosen correctly.

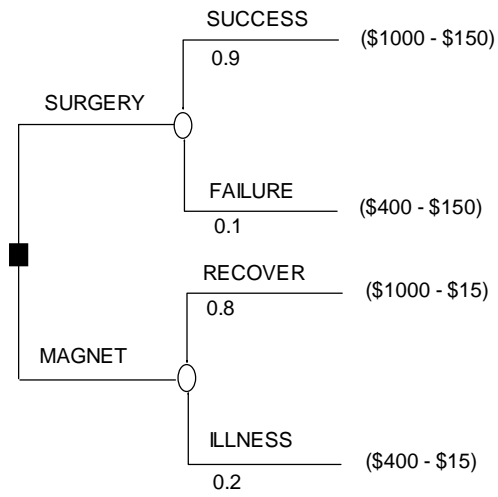


Figure 61: Example of a decision tree for traumatic reticulitis treatment

Epidemiological Animal Disease Information Management

Learning Objectives

At the completion of this topic, you will be able to:

- enumerate steps to take during an outbreak investigation, including description of the outbreak by animal, place and time
- understand the principles of herd health and productivity profiling

Outbreak Investigation

An outbreak is a series of disease events clustered in time. Outbreak investigations are used to systematically identify causes (risk factors) of disease outbreaks or to identify patterns of disease occurrence. The investigator asks the questions: What is the problem? Can something be done to control it? Can future occurrences be prevented? The problem can be approached using the traditional *clinical* or the modern *epidemiological* approach to investigation of herd problems.

Clinical approach

With the clinical approach, the focus is on diseased animals and they are identified on the basis of features which distinguish them from normal animals. The clinician can come up with a correct diagnosis given that the “true” diagnosis is included in the list of differential diagnoses taken into consideration. The latter will depend on the areas of special expertise, and typically clinicians will only include differential diagnoses they are familiar with. The clinical diagnosis can be derived relatively easily if it is a single clearly identifiable disease. The situation becomes more difficult, if multiple disease determinants interact to cause a syndrome. This type of outbreak situation is better described as a causal web than a causal chain.

Epidemiological approach

In contrast, the epidemiological approach to investigation of herd problems removes the assumption of the existence of a ‘normal’ standard. It focuses on the comparison of sub-groups or animals. A systematic approach is used to keep the process objective and unbiased. With simple problems, the epidemiological will not be distinguishable from the clinical approach. In the case of new or complex problems, the epidemiological approach becomes the method of choice for outbreak investigations.

The *epidemiological approach* to investigations of herd problems includes the following series of investigational steps: First the *diagnosis* has to be verified by making a definitive or tentative diagnosis, followed by a clinico-pathological work-up. Then a *case definition* is established. Cases should be defined as precisely as possible, and other diseases should be excluded. As the next step, the *magnitude of the problem* has to be determined. Is there an epidemic? The cumulative incidence can be computed and compared with normal or expected risks of disease. Subsequently, the *temporal pattern* is examined which typically involves constructing an epidemic curve. If possible, incubation and exposure period are estimated on the basis of the data. Then, the *spatial pattern* is examined, for example by drawing a sketch map of the area or the layout of the pens and the number of cases within pens. The investigator should inspect the drawing for possible interrelationships among cases, and between location and cases and other physical features. This process is followed by an examination of *animal patterns*. It includes investigating factors such as age, sex, breed and strain patterns. A list of potential causal or non-causal factors associated with the disease is established. Animals are categorised according to the presence of each attribute. This data can be presented as frequency and attack rate tables. After collating this information it should be analysed using quantitative methods. As part of the *data analysis*, factor-specific disease risks are computed for each of the potential

risk factors. The objective is to identify highest as well as lowest disease risks, the greatest difference between disease risks, to estimate the relative and attributable risks. Information about the expected level of disease is very important in this process. The data should not just be looked at by calculating proportions, it is also important to assess absolute numbers of animals. Risk factors associated with the epidemic are identified by assessing the association between disease patterns and the distribution of factors. The objective is to demonstrate that an observed association is not due to chance. The result from this analysis should be a working hypothesis taking into account potential causes, sources, mode of transmission, exposure period and population at risk. The hypothesis may have to be revised, if it does not fit all facts. If it is possible, the hypothesis should be tested. With complex problems not revealing any quick answers as to the source of the problem, it is often advised to conduct an intensive follow-up. This would involve a clinical, pathological, microbiological and toxicological examination of tissues, feeds, objects etc.. Detailed diagrams of feed preparation or movement of animals could be prepared and a search for additional cases on other premises or outbreaks of similar nature in other locations could be conducted. If it is considered desirable to test the hypotheses, the investigators could conduct an intervention study. The outcome of the outbreak investigation should be presented as a written report containing recommendations for controlling the outbreak and preventing further occurrences.

Example of an outbreak investigation

(adapted from Gardner, I. 1990: Case study: Investigating neo-natal diarrhoea. In D. Kennedy (editor) Epidemiology at Work. Refresher Course for Veterinarians. Proceedings 144. Postgraduate Committee in Veterinary Science, University of Sydney, Sydney, Australia. 109-129.)

To illustrate the above concepts the example of an investigation of an outbreak of neonatal diarrhoea in pigs will be used. The *history* of the outbreak is that the farmer reported an ongoing diarrhoea problem in neonatal pigs in his 150-sow breeding/finishing herd. During

the 12 months prior to the outbreak typically about 7% of the litters suffered from diarrhoea, now 40% of litters are affected. The immediate action taken by the vet included submission of 3 acutely affected pigs to the local diagnostic laboratory who diagnosed one piglet as infected with *E.coli* serotype 08, but did not find any pathogenic bacteria or viruses in the other 2 pigs. The intestinal lesions found in all 3 pigs were consistent with acute enteritis. The veterinarian decides to adopt an epidemiological approach for investigating the problem and its financial impact.

General knowledge about preweaning diarrhoea in pigs indicates that most herds have endemic neonatal diarrhoea at low levels and experience periodic outbreaks. The organisms involved are *E.coli*, rota virus, *Salmonella spp.*, *Campylobacter spp.*, cryptosporidia and coccidia, as well as TGE virus in some countries. There is still insufficient knowledge to explain why some litters (herds) are affected and others not. Figure 62 presents the causal web of factors associated with the occurrence of coli-bacillosis in neonatal pigs.

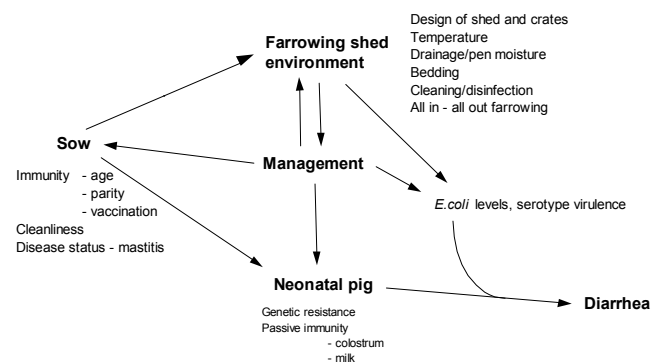


Figure 62: Multi-factorial web of causal factors involved in piglet diarrhea

In this particular case, as a first step in the investigation the diagnosis has to be verified. For this purpose, further dead pigs are being necropsied and more samples submitted to the diagnostic laboratory. Rectal swabs are taken from scouring and non-scouring piglets and also sent to the laboratory. A case definition for deaths due to scouring is based on the following signs: External and internal evidence of diarrhoea, signs of dehydration,

intestinal contents abnormally fluid-like and staining of peri-anal region. The magnitude of the problem is quantified as follows:

- 11 (42.3%) of 26 litters are affected (in this herd 7% and in the industry 5% are considered normal)
- scours-specific risk:

$$\frac{\text{\#deaths due to scours}}{\text{\# pigs born alive}} = \frac{24}{253} = 0.09$$
- proportional mortality risk for scours:

$$\frac{\text{\#deaths due to scours}}{\text{\# total dead pigs}} = \frac{24}{48} = 0.50$$

As a next step the temporal pattern is defined which typically involves constructing an epidemic curve based on the number of litters affected per week (see Figure 63). In this particular case, the resulting curve is difficult to interpret, because the numbers of litters are too small.

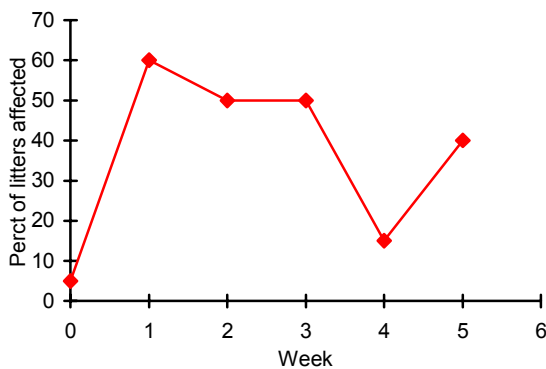


Figure 63: Temporal pattern of neonatal diarrhoea

The spatial pattern is investigated by drawing a sketch map of the layout of the farrowing house (see Figure 64). The result suggests that the pens on the western side of the shed have more affected litters farrowed per pen.

PEN #	LITTERS AFFECTED
1	1
2	1
3	1
4	2
5	0
6	1
7	2
8	1
9	1
10	0
11	0
12	0
13	1
14	0
15	0
16	0

west (near entrance)
 ↑
 ↓
 east, near extraction fans

Figure 64: Spatial pattern of neonatal diarrhoea

One factor related to the animal pattern is the age at which different litters were treated in the context of all litters at risk (see Figure 65). The data suggests that most litters with diarrhoea were either 3 or 5 days old, when they were treated. But the data does not indicate any particular age pattern within this range of days.

age at treatment (days)	LITTERS	
	affected	total
3	3	8
4	1	8
5	3	8
6	1	8

Figure 65: Litters affected by age of treatment

As a next step the data is analysed using cross-tabulations to evaluate the effects of parity, litter size, whether the sow was sick at farrowing and location of crate on the risk of a litter being affected. Relative risks are calculated and chi-square tests are used to assess their statistical significance. The first factor analysed is the parity distribution of the sows (see Figure 66). The chi-square test is used to assess if there is an association between deaths due to scours in piglets and parity of sow. The resulting p-value is 0.34, which indicates that given the sample size there is no statistical association between

parity and risk of scouring pigs in a litter. The problem with this particular application of the chi-square statistic is that the expected number of counts in at least one cell is less than 5 observations, which violates the assumptions of this statistical method. Modern statistical software can get around this problem by using Monte Carlo sampling or exact inference techniques. Applying the Monte Carlo method results in a significance level of 0.39 (95% CI 0.38 - 0.40). It should be kept in mind, that due to the small sample size the statistical power of this analysis is low.

Parity	Litters with dead scouring pigs			
	Proportion	yes	no	total
1	0.80	4	1	5
2	0.60	3	2	5
3	0.25	1	3	4
4	0.50	2	2	4
5	0.33	1	2	3
6	0.00	0	3	3
7	0.00	0	1	1
8	0.00	0	1	1
Totals	0.42	11	15	26

$$\chi^2 = 7.9 ; 7 \text{ df} ; p = 0.34$$

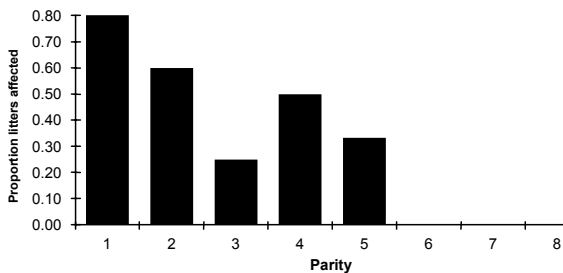


Figure 66: Example calculations and bar chart of the risk of a litter to be affected by neonatal diarrhoea stratified by parity

Visual inspection of the data taking into account potential biological mechanisms suggests that it might be useful to focus the analysis on comparing litters from parity 1 with those from older sows. The results of this analysis are presented in Table 8. The relative risk calculation suggests that first parity sows were 2.4 times as likely to have dead scouring pigs in their litter as older sows. The chi-square value has a p-value of 0.06, but has to be interpreted with caution, because the expected number of counts in two cells is less than 5. The exact inference indicates a p-value of 0.13 which is not significant. The next

variable to look at is the disease status of the sow (see Table 9). The relative risk amounts to 2.7, but also is not statistically significant.

Table 8: Cross-tabulation of aggregated parity categories against presence of dead scouring pigs in litter

parity	dead scouring pigs in litter			total
	proportion	yes	no	
1	0.80	4	1	5
> 1	0.33	7	14	21
totals	0.42	11	15	26

$$RR = \frac{4/5}{7/21} = 2.4$$

$$\chi^2 = 3.6; 1df; p = 0.06$$

Fisher's exact test (2-tailed) p=0.13

Table 9: Cross-tabulation of disease status of sow against presence of dead scouring pigs in litter

sick sow	dead scouring pigs in litter			total
	proportion	yes	no	
yes	1.00	2	0	2
no	0.38	9	15	24
totals	0.46	11	15	24

$$RR = \frac{2/2}{9/24} = 2.67$$

$$\chi^2 = 3.0; 1df; p = 0.09$$

Fisher's exact test (2-tailed) p=0.17

The association between litter size groupings and presence of dead scouring pigs in the litter does not show statistical significance (see Table 10). On the other hand there is a statistically significant association between pens being in the part of the shed close to the entrance or not and presence of dead scouring pigs in the litter (see Table 11). Litters farrowed near the entrance were 3.86 times as likely as litters near the extraction fans to have dead scouring pigs.

Table 10: Cross-tabulation of litter size categories against presence of dead scouring pigs in litter

litter size	dead scouring pigs in litter			total	RR
	proportion	yes	no		
<=8	0.13	1	7	8	1
9 - 10	0.50	4	4	8	4
11 - 12	0.50	4	4	8	4
>=13	1.00	2	0	2	8
totals	0.42	11	15	26	

$$\chi^2 = 6.0; 3df; p = 0.11$$

Table 11: Cross-tabulation of pen location categories against presence of dead scouring pigs in litter

pens	dead scouring pigs in litter			total
	proportion	yes	no	
1 - 8	0.75	9	5	14
9 - 16	0.17	2	10	12
totals	0.42	11	15	26

$$RR = \frac{9/14}{2/12} = 3.86$$

$$\chi^2 = 6.0; 1df; p = 0.01$$

Fisher's exact test (2-tailed) p=0.02

From this first stage of the epidemiological analysis it can be concluded that there are potentially multiple factors associated with the disease problem. In this type of situation, it is then necessary to decide on which of these factors are most important and which are likely to be confounding factors such as could be for example gilts which may be more likely to be placed near the entrance. Statistical analysis methods for multivariable problems are appropriate with this type of data. In this case, *logistic regression* is the method of choice as it allows multivariate analysis of binary outcome variables. The outcome variable is whether *E. coli* diarrhoea was *present* or *not present* in a litter. The result of the logistic regression analysis indicates that *parity number* (PARITY) is the most important risk factor, but litter size (BORN) and crate location (CRATENO) were also included (see Figure 67). There was no statistically significant interaction between *crate* and *parity number* which suggests that the two risk factors are not dependent. It is also unlikely that the two factors were

confounded, as removal of each from the model did not change the magnitude of the effect of the other factor to a meaningful extent. *Disease status of the sow* was not important as it did not produce a statistically significant regression coefficient. The results can be interpreted as suggesting that the risk of diarrhoea diminishes with parity and with distance from the entrance to the shed. The odds ratio of 0.34 for PARITY indicates that an increase in one unit in the variable PARITY changes the odds of mortality from scouring in the litter by a factor of 0.34. An increase in litter size results in an increased risk of mortality from scouring in the litter.

```

STATISTIX 4.0                                COLIPIGS, 25/03/93, 8:29

UNWEIGHTED LOGISTIC REGRESSION OF COLI

PREDICTOR
VARIABLES  COEFFICIENT  STD ERROR  COEF/SE  P
-----
CONSTANT   -5.89972    4.49999   -1.31     0.1898
PARITY     -1.06609    0.51580   -2.07     0.0387
BORN       1.30874    0.63633    2.06     0.0397
CRATENO    -0.42139    0.22000   -1.92     0.0554

DEVIANCE                                12.21
P-VALUE                                  0.9530
DEGREES OF FREEDOM                       22

CASES INCLUDED 26  MISSING CASES 0
    
```

```

STATISTIX 4.0                                COLIPIGS, 25/03/93, 8:29

LOGISTIC REGRESSION ODDS RATIOS FOR COLI

PREDICTOR    95% C.I.
VARIABLES   LOWER LIMIT  ODDS RATIO  UPPER LIMIT
-----
PARITY       0.13         0.34        0.95
CRATENO      0.43         0.66        1.01
BORN         1.06         3.70       12.88
    
```

Figure 67: Output generated by multiple logistic regression analysis of piglet scouring data

The results of this data analysis have to be interpreted very carefully. Firstly, in this particular case the statistical power was quite low due to the small number of 26 litters available for analysis. Under such circumstances, the investigator could use a 10% significance level, but should be aware that this will increase the risk of detecting statistically significant associations (=type I error) which in reality might not exist. With any observational study one should be aware of the potential for confounding factors. In this particular data set, the gilts not being familiar with the farrowing facilities may

have been more likely to be placed in pens close to the entrance to the farrowing shed. The potential for such a confounding relationship between parity category and crate location was eliminated as part of the multivariate analysis. Taking the results together, risk of diarrhoea does seem to be related to parity and litter size which could both indicate insufficient passive immunity. The variable crate location may be an indicator for inadequate disinfection and cleaning procedures. Hence, there are two main hypotheses which could be further investigated to come up with a set of suitable preventive measures. To test the hypothesis of poor passive immunity, the investigator could consider conducting an intervention study of sow vaccination (*E.coli*).

From the farmer's perspective, an economic evaluation will be important. The preweaning mortality is currently at 19%, but used to be 11.5% which means that 7.5% are attributable to the current scouring episode. This excess mortality of 7.5% in a total of 253 pigs born alive results in 19 piglet deaths, of which 1 (5%) would be expected to die after weaning, leaving 18 pigs for the remaining calculations. During the last year 7% of litters were treated, meaning that 1.8 litters (7%) of the current 26 litters would be expected to be treated under normal circumstances, but in reality 8 (31%) of the 26 litters had to be treated in association with the outbreak. This results in opportunity costs for 18 pigs of $18 * \$35$, treatment costs for 6 litters of $6 * \$10$ amounting to a total of \$690 which comes to about \$26.50 per sow farrowed. Sow vaccination costs are estimated as \$5 for the vaccine plus \$0.30 for labour resulting in a total of \$5.30. The benefit/cost ratio is therefore $26.5 / 5.3 = 5$. This means introduction of vaccination could very beneficial. For every dollar spent, 5 dollars could be earned.

As a consequence of this analysis, the following action could be taken. Firstly, the intervention study of sow vaccination with *E.coli* vaccine mentioned above could be conducted. The spatial pattern could be further investigated by reviewing the cleaning

and disinfection program on the farm. Additional samples for microbiology could be collected from untreated pigs with diarrhoea and pigs without diarrhoea. The investigator should remember that with this type of problem it is rather unlikely to have only one causal agent involved and that culture has poor sensitivity. This means for the laboratory not to isolate any other potentially causative organisms does not necessarily mean that they have not been present. There is also the possibility that the herd has a problem with post-farrowing illness in sows, but given the small number of only 2 cases it is unlikely to have a major impact in the current problem. In any case, it would be recommended to monitor piglet mortality over a period of time. This will also allow review of the hypotheses as more data becomes available. Finally, the investigator should produce a report for the farmer describing the outcome of the investigation.

Assessment of Productivity and Health Status of Livestock Populations

Traditionally, animal disease control mainly focused on monitoring of disease outbreaks and movements of disease agents. Nowadays, it has become important to collect information for setting priorities and defining actions against diseases common in livestock population. *Health and productivity profiling* is one particular method that can be used for this objective. With this method of data collection, a limited number of representative sample units such as for example a small number of herds or villages are being used to collect detailed information on production, management and disease. This data forms the basis of epidemiological analysis for investigation of complex relationships including interactions between multiple factors and diseases. The results can be applied to defining action priorities for disease control programmes based on economic or other considerations.

In developing countries an effective government veterinary service often is not economically sustainable, but can be supplemented by a *primary animal health*

care service which employs so called "barefoot" veterinarians to provide the most common veterinary services to farmers. These priority services can be defined on the basis of a basic set of health and productivity indicators which are being continually monitored using *village animal health and production monitoring systems*.

Theoretical Epidemiology

Simulation modelling is an area which will become more important as part of decision support in agriculture in the future. Models are used to represent dynamic processes or systems and simulate their behaviour through time. They mimic the system under study which allows testing of specific hypotheses for example about the epidemiology of an infectious process or to identify gaps in our understanding which need further investigation. As part of a decision support system they can be used to test alternative strategies. There are two main groups of simulation models, those based on mathematical equations which are called *deterministic models* and those based on probabilistic sampling of distributions which are called *stochastic models*. The tools used for the development of such models include computerised spreadsheets which allow non-programmers to construct simple models. Computer programming languages have to be used when developing more complex models.

Information Technology and Veterinary Applications

Modern information technology will in the near future become part of the tool set which the veterinarian in the 20th century will have to work with. *Computerised information retrieval systems* have been developed such as for example the electronic edition of the MERCK VETERINARY MANUAL. Electronic literature reference databases such as Current Contents can now be accessed through the World Wide Web. On the more technical side, artificial intelligence methods including knowledge-based systems (expert systems) or neural networks are being used to design decision support systems which can help the veterinarian in the diagnostic process.

Software applications used for individual cases apply a probabilistic approach to veterinary diagnosis and therapy. Examples of such systems are BOVID and CANID which are expert systems used for diagnoses of cattle and canine disease problems, respectively. Herd health management increasingly involves the use of computerised recording and monitoring systems such as DairyWIN or PigWIN. Eventually these systems will develop into decision support systems (see Figure 68).

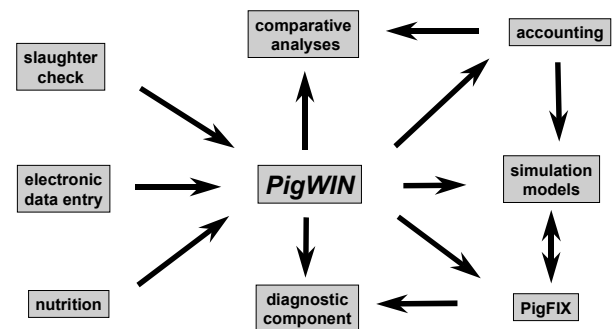


Figure 68: Components of a decision support system for pig production

Modern national disease control programs have computerised database management systems as an essential component. Examples are the New Zealand national tuberculosis database and the ANIMO system used by the European Union to monitor animal movements. More recently, geographical information system technology has developed to a level where it can be used by disease control managers at a farm as well as at the regional level (see Figure 69). For these systems to evolve into a decision support system they should integrate database management, simulation modelling, decision analysis as well as expert system components. The computerised emergency response system for outbreaks of foot-and-mouth-disease called EpiMAN developed in New Zealand is one of the most sophisticated such systems currently available (see Figure 70).

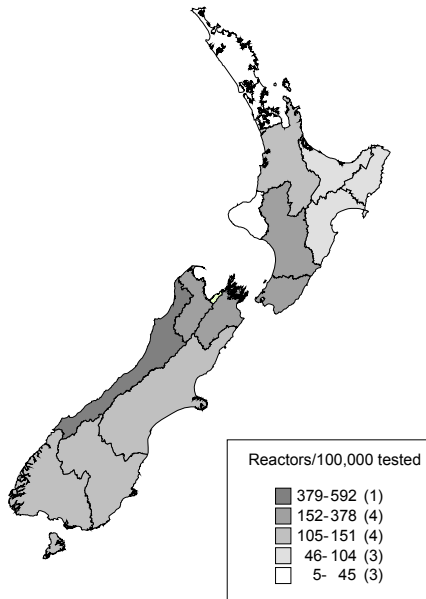


Figure 69: Choropleth map of reactors to the tuberculin test in cattle in New Zealand during 1994/95

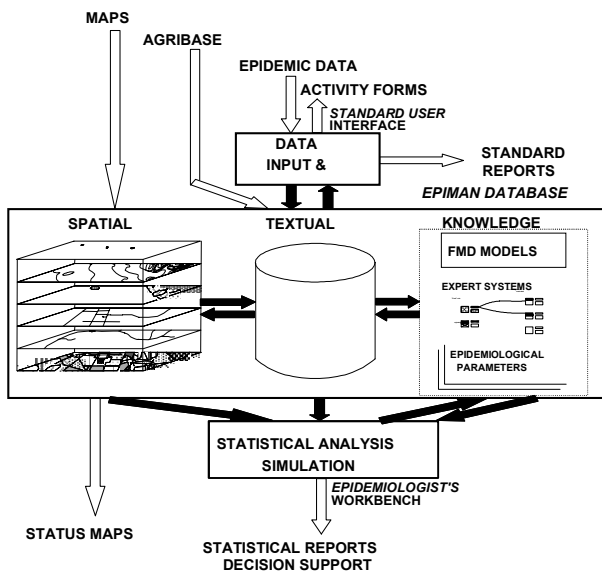


Figure 70: Schematic diagram of EpiMAN decision support system for foot-and-mouth disease control

Recommended Reading

Dawson,B. and Trapp,R.G. 2001: *Basic and clinical biostatistics*. 3rd ed. Lange Medical Books/McGraw-Hill, New York, 399pp.

(excellent and very readable introduction to applied biomedical statistics)

Fletcher,R.H., Fletcher,S.W. and Wagner,E.H. 1996: *Clinical epidemiology*. 3rd ed., Williams & Wilkins, Baltimore, U.S.A., 276pp.

Hulley,S.B., Cummings,S.R., Browner,W.S., Grady,D., Hearst,N. and Newman,T.B. 2001: *Designing clinical research: An epidemiologic approach*. 2nd ed., Lippincott Williams & Wilkins, Philadelphia. 336pp.

(very good text on study design)

Martin,S.W., Meek,A.H. and Willeberg,P. 1987: *Veterinary epidemiology*. Iowa State University Press, Ames, Iowa, U.S.A., 343pp.

(recommended as a general reference)

Noordhuizen,J.P.T.M., Thrusfield,M.V., Frankena,K. and Graat,E.A.M. 2001: *Application of quantitative methods in veterinary epidemiology*. 2nd reprint. Wageningen Pers, Wageningen, The Netherlands, 430pp.

(covers basic as well as more advanced veterinary epidemiology)

Rothman,K.J. and Greenland,S. 1998: *Modern epidemiology*. 2nd ed., Lippincott - Raven, Philadelphia, U.S.A., 737pp.

(very detailed text book covering basic as well as advanced epidemiological analyses)

Smith,R.D. 1995: *Veterinary clinical epidemiology - A problem-oriented approach*. 2nd ed., CRC Press, Boca Raton, Florida, 279pp.

(recommended as a reference for clinical veterinary epidemiology)

Thrusfield,M. 1995: *Veterinary epidemiology*. 2nd ed., Blackwell Science, Oxford, England, 479pp.

(recommended as a general veterinary epidemiology reference)

Toma,B., Dufour,B., Sanaa,M., Benet,J.-J., Ellis,P., Moutou,F. and Louza,A. 1999: *Applied veterinary epidemiology and the control of disease in populations*. Maisons-Alfort, France: AEEMA, 536pp.

(recommended as a general veterinary epidemiology reference)

Index

—A—

accuracy, 39
adjusted risk estimates, 16
agreement, 40
analytical epidemiology, 18
animal pattern, 53
antigenic variation, 8
apparent prevalence, 41
approximate relative risk, 23
artificial intelligence, 59
attack rate, 15
attributable fraction, 23
attributable risk, 23

—B—

basic epidemiological concepts, 5
batteries of multiple tests, 50
bias, 25, 39

—C—

carrier state, 8
case fatality rate, 15
case series, 19
case-control study, 20
causation, 7, 9
cause-specific mortality rate, 15
censoring, 15
census, 19
central limit theorem, 35
challenges, 3
chance, 25
cluster sampling, 30, 31
comparison of diagnostic tests, 39
confidence intervals, 29
confounding, 26
contact transmission, 8
convenience sampling, 29
cross-product ratio, 23
cross-sectional study, 21
crude mortality rate, 15
cumulative incidence, 12
cumulative incidence difference, 23
cumulative incidence ratio, 22

—D—

decision analysis, 51
decision support system, 59
descriptive epidemiology, 12
determinants of disease, 7
diagnostic test, 39
diagnostic test performance, 39
diagnostic testing for disease control and eradication, 49
diagnostic tests, 38
diagnostic value method, 44
direct standardisation, 16
disease, 5
disease vectors, 8

—E—

epidemiological studies, 19
epidemiological triad, 10

establishing cause, 11
etiologic fraction, 23
evaluation of diagnostic tests, 39
Evan's unified concept of causation, 9
evidence-based veterinary medicine, 4
excess risk, 23
experimental studies, 19
exposure, 22
extrinsic determinants of disease, 9

—F—

force of morbidity, 13
force of mortality, 13

—G—

gaussian distribution method, 44
gold standard, 39

—H—

hazard function, 15
hazard rate, 13
health and productivity profiling, 58
Henle - Koch postulates, 9
herd health management software, 59
host/agent relationship, 8

—I—

incidence density, 13
incidence rate ratio, 23
incidence risk, 12
incubation period, 8
indirect standardisation, 17
infection, 7
infectivity, 7
information technology, 59
interaction, 26
intermediate hosts, 8
intrinsic factors, 7
intrinsic host determinants, 8

—J—

judgmental sampling, 29

—K—

kappa test, 40

—L—

level of disease occurrence, 35
likelihood ratio, 45
longitudinal studies, 19

—M—

measurement of disease frequency, 12
methods of transmission, 8
misclassification, 25
mortality rate, 15
multistage sampling, 30, 32

—N—

necessary cause, 10
non-observational studies, 19
non-probability sampling, 29
normal/abnormal criteria, 44

—O—

observational studies, 19, 20
odds ratio, 23
outbreak investigation, 53

—P—

p - value, 25
pathogenicity, 7
percentile method, 44
period of communicability, 8
population, 5
potential impact, 22
power, 25
precision, 39
predictive value, 40, 43
predictive value method, 44
prepatent period, 8
pre-test probability, 42
prevalence, 13
prevalence and diagnostic tests, 41
prevalence difference, 23
prevalence ratio, 22
prevented fraction, 24
probability of not detecting disease, 36
probability sampling, 29
prospective cohort study, 20
prospective studies, 19
purposive sampling, 29

—R—

rate ratio, 23
relative odds, 23
relative risk, 22
retrospective studies, 19
risk, 22
risk assessment, 22
risk difference, 23
risk factor method, 44
risk factors, 22
risk ratio, 22
ROC Analysis, 45

—S—

sample, 19
sampling, 28
sampling fraction, 28
sampling frame, 28
sampling to detect disease, 36, 37
sampling to measure continuous variable, 37
sampling unit, 28
selection bias, 25
sensitivity, 39
simple random sampling, 29, 30
simulation modelling, 59
sources of error in probability sampling, 29
spatial pattern, 7, 53
specificity, 40

—β—

β - error, 25

—S—

standardisation of risk, 16
standardised risk estimates, 16
strategies for diagnostic test selection, 42
stratified sampling, 30
strength of association, 22
study, 18
study population, 28
sufficient cause, 10
survey, 18
survival, 15
survival function, 15
systematic sampling, 29, 30

—T—

target population, 28
temporal pattern, 5, 53
test interpretation at the individual level, 40
test performance, 40
theoretical epidemiology, 59
therapeutic method, 44
true incidence rate, 13
true prevalence, 41
type I error, 25
type II error, 25

—V—

vaccine efficacy, 24
vehicular transmission, 8
virulence, 8

—W—

web of causation, 9

—□—

α- error, 25