



**AgEcon** SEARCH  
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

*The World's Largest Open Access Agricultural & Applied Economics Digital Library*

**This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.**

**Help ensure our sustainability.**

Give to AgEcon Search

AgEcon Search

<http://ageconsearch.umn.edu>

[aesearch@umn.edu](mailto:aesearch@umn.edu)

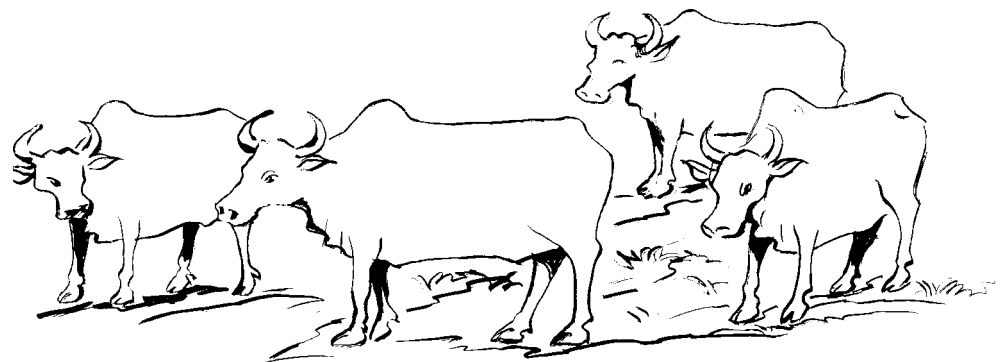
*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

*No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.*

# **Survey Toolbox for Livestock Diseases**

**A Practical Manual and Software Package for  
Active Surveillance in Developing Countries**

**Angus Cameron**



The Australian Centre for International Agricultural Research (ACIAR) was established in June 1982 by an Act of the Australian Parliament. Its mandate is to help identify agricultural problems in developing countries and to commission collaborative research between Australian and developing country researchers in fields where Australia has a special research competence.

Where trade names are used this constitutes neither endorsement of nor discrimination against any product by the Centre.

© Australian Centre for International Agricultural Research, 1999

Cameron, A.R. 1999. Survey Toolbox for Livestock Diseases - A Practical Manual and Software Package for Active Surveillance in Developing Countries  
ACIAR Monograph N° 54, vii + 330 p.

ISBN 1 86320 234 X

Illustrations: Kongphat Luangrath

# Acknowledgments

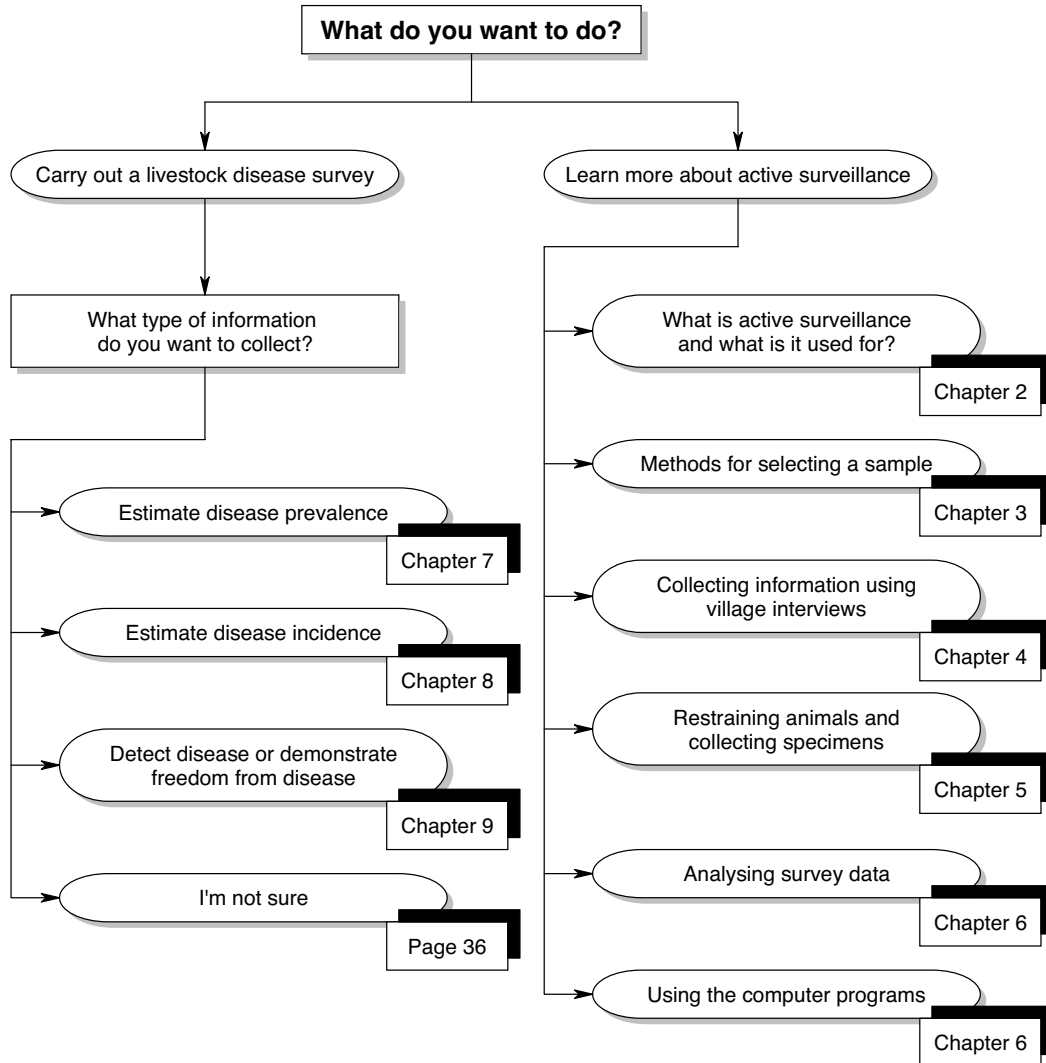
The surveillance techniques, procedures and software described in this book were developed during the course of two research projects funded by the Australian Centre for International Agricultural Research (ACIAR). The first was entitled “Improved methods in diagnosis, epidemiology, economics and information management in Australia and Thailand”, and ran from 1994 to 1996. The second, “Development of field survey and information management techniques for animal health priority setting in the Lao People’s Democratic Republic” ran from 1996 to 1998.

I would like to acknowledge the significant contributions of all who have been involved in the projects.

- ACIAR, and in particular, Dr John Copland
- Thai Department of Livestock Development: Dr Pornchai Chamnanpood, Head of Epidemiology Section, Northern Veterinary Research and Diagnostic Laboratory, and staff
- Lao Department of Livestock and Fisheries: Mr Singkham Phonvisay, Director General of the Department; Dr Mahanakhone Sourinya, Head of Animal Health Division; Dr Sounthone Vonthilath, Deputy Head; Dr Siseng Khounsy, Head of Research Section; and staff
- University of Queensland: Dr Pramod Sharma, Department of Geographical Sciences and Planning; Dr Steve Harrison, Department of Economics
- Australian Animal Health Laboratory (AAHL): Dr Harvey Westbury and Mr Stuart Blacksell
- Staff of the Queensland Department of Primary Industries
- Members of the veterinary services of Philippines, Vietnam, Cambodia, and Malaysia
- Dr Murray Maclean of the Cambodia-Australia Agricultural Extension Project
- The smallholder livestock owners of Laos and northern Thailand, who participated in surveillance activities
- and especially Dr Chris Baldock, of AusVet Animal Health Services.

For Catriona

# Guide to the Manual



# Contents

Chapter 1:	Introduction .....	1
<b>Part I: Background to Disease Surveys .....</b>		<b>11</b>
Chapter 2:	General Principles of Animal Disease Surveillance .....	13
Chapter 3:	Sampling .....	37
Chapter 4:	Principles of Data and Specimen Collection .....	83
Chapter 5:	Village Interviews .....	103
Chapter 6:	Computerised Data Management and Analysis .....	123
<b>Part II: Survey Design and Analysis .....</b>		<b>143</b>
Chapter 7:	Prevalence Surveys .....	149
Chapter 8:	Incidence Rate Surveys .....	169
Chapter 9:	Surveys to Demonstrate Freedom from Disease .....	189
<b>Part III: Notes for Trainers .....</b>		<b>209</b>
Chapter 10:	Guide for Trainers .....	211
Chapter 11:	Lesson Plans .....	221
Chapter 12:	Activity Sheets .....	257
<b>Part IV: Appendices .....</b>		<b>283</b>
Appendix A:	Glossary of Epidemiological Terms .....	285
Appendix B:	Statistical Equations .....	295
Appendix C:	Computer Programs .....	309
Appendix D:	Example Data Collection Forms .....	317
Appendix E:	Contents of CD .....	325
<b>Index .....</b>		<b>327</b>

# 1

## Introduction

### Contents

Chapter 1: Introduction  
Active surveillance  
Who is this book for?  
How to use this book and software

#### **Part I: Background to Disease Surveys**

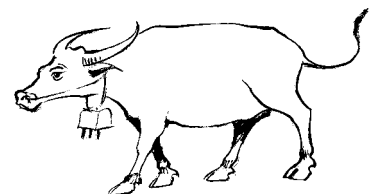
Chapter 2: General Principles of Animal Disease Surveillance  
Chapter 3: Sampling  
Chapter 4: Principles of Data and Specimen Collection  
Chapter 5: Village Interviews  
Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys  
Chapter 8: Incidence Rate Surveys  
Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers  
Chapter 11: Lesson Plans  
Chapter 12: Activity Sheets



## Importance of livestock diseases

Livestock diseases have a significant impact in many developing countries. Most people in many of these countries live in rural areas, and depend on agriculture for their livelihoods. Raising livestock may provide a good source of food through meat, milk and eggs; manure for fertiliser and fuel; fleece and hides for clothing; and draught power for working fields and transport. Livestock may also be a form of savings, which can be cashed in through the sale of animals when needed. In addition to the benefits to individual livestock owners, both smallholder and intensive livestock industries may make a large contribution to the national economy, particularly in developing countries with few other major industries.

Diseases of livestock can cause major losses, both to livestock owners and the country as a whole. In many developing countries, outbreaks of major diseases occur frequently, and are poorly controlled, resulting in large numbers of deaths. Less spectacular chronic or subclinical diseases may also cause high losses through decreased fertility, decreased weight-gain, inefficient use of feed or the inability of animals to work. Zoonotic diseases (diseases affecting both animals and humans) may have an important impact on public health.

The veterinary authorities of many developing countries are faced with a quandary. The livestock disease problems they face are much greater and cause higher losses than those faced by their counterparts in developed countries, but the resources available to address them are often inadequate.

Controlling major livestock diseases in largely rural societies has the potential to dramatically improve the quality of life of rural poor. Financial and other resource are severely limited, so disease control programs in developing countries must use available resources as efficiently as possible. Successful techniques for disease control which have been used in developed countries are often too expensive or impractical for use in developing countries.

There is therefore a need for practical, efficient techniques for disease control which have been specifically developed to take account of the constraints faced by developing countries. This book aims to address some of those needs through improved information collection.

## Strategic disease control

One way to control disease more efficiently is to use a “smart” or strategic approach. For instance, rather than using an expensive “blanket” approach to vaccination in which an attempt is made to vaccinate every animal, an understanding of the spread of the disease in the population may allow a much less expensive approach to be used.

**Example:** Imagine a country with a centrally located intensive cattle growing area. This area sends cattle to less intensive smallholders in all other parts of the country. If the level of disease in this area is high, it will be spread through the whole country. Targeted vaccination of animals in this small intensive area may have a significant impact on the number of cases in other parts of the country, as this is the source of most infection.

While the pattern of disease is rarely as simple as presented in this example, it demonstrates one important point. In order to control disease efficiently, it is first necessary to understand the distribution of disease. In this example, we needed to know 1) where the intensive and non-intensive cattle industries were located, 2) the level of disease in the different areas, and 3) the pattern of livestock movements between different areas. Given this information, we can develop a more efficient control strategy – targeted vaccination of the intensive area only. This would involve

vaccinating a smaller number of animals, as well as lower costs for vaccine distribution (as transport infrastructure is likely to be more developed in an intensive area). This strategy may well be able to bring the level of disease throughout the country to a low enough level so that it is either no longer causing significant losses, or it can be completely eradicated through stamping out (slaughtering all animals involved in an outbreak).

This book deals with a series of techniques for the collection of animal health information in developing countries. The techniques have been developed to provide high quality information at a minimal cost, taking infrastructure and other constraints into account. The information collected will provide a clear understanding of disease patterns and, in turn, can be used to develop smart, efficient disease control programs, as well as for other purposes discussed in Chapter 2.

The information is used for monitoring and surveillance. The world organisation for animal health, OIE (Office International des Epizooties) defines these terms as follows:

Surveillance	Surveillance means the continuous investigation of a given population to detect the occurrence of disease for control purposes, which may involve testing of a part of the population.
Monitoring	Monitoring constitutes on-going programmes directed at the detection of changes in the prevalence of disease in a given population and in its environment.

## Active surveillance

Active surveillance

At the heart of the techniques used in this book is the use of active surveillance. As discussed in Chapter 2, traditional disease reporting systems, such as compulsory disease outbreak notifications, or the use of laboratory submission data (known as passive surveillance) have a number of disadvantages. Under-reporting, expense and non-representative reports are some of the main problems. Active surveillance differs from a passive reporting system in that it uses surveys of a relatively small, representative sample of the population to gather specific information about that population. The key advantages of active surveillance are that the quality of information collected is usually better, the information reflects the true situation in the entire population, and it is often faster and cheaper to collect than with passive methods.

Passive surveillance

Despite their problems, passive reporting systems are an important source of disease information. Passive disease reporting systems in one form or other are in place in virtually all countries, but relatively few countries make regular use of active surveillance, in spite of the advantages. This is partly because appropriate techniques were not previously available, and veterinary staff were not trained in the skills necessary. This book describes how appropriate active surveillance techniques can be implemented and provides the necessary skills.

## Who is this book for?

This book is for people working in the area of livestock disease control in developing countries, primarily, but not exclusively, as part of the government veterinary services. It is designed to be used by three different levels of staff.

National staff

Firstly, it is aimed at national or sub-national epidemiologists, responsible for the planning and monitoring of disease control programs and the collection and analysis of animal health information. Included in this group are those working for international aid or development organisations. The book provides all the information and tools necessary to plan, implement, and analyse livestock disease surveys, as well as assistance with training field staff.

Provincial staff

The second group for whom this book will be useful is state or provincial veterinary staff responsible for the implementation of disease surveys. The book provides a detailed description of the field techniques necessary to successfully carry out a disease survey. This group may require a less detailed understanding of survey design.

Local or district staff

Thirdly, local or district veterinary staff, responsible for much of the field work involved in disease surveys will also find all the procedures and techniques required in this book. These staff do not need a thorough understanding of survey design and analysis techniques, but need a good understanding of specimen collection, random sampling and village interview techniques.

## How to use this book and software

Survey Toolbox

This book is accompanied by a set of software programs, collectively called Survey Toolbox. The book and software are closely integrated and are designed to be flexible, depending on the needs and preferences of the user.

### Using the software

The purpose of the computer programs is to assist with the planning, implementation, and analysis of the surveys described in the book. Many survey procedures (such as random selection from a sampling frame, or the analysis of data collected from a two-stage seroprevalence survey) are either time consuming or require extremely complex formulae and specialist statistical skills. In order to allow veterinary staff to carry out and analyse surveys without expert statistical assistance, all these procedures and formulae have been implemented in a set of computer programs, specifically designed for the surveys described in this book.

Each program performs a particular task and can be used independently. The programs do *not* provide a comprehensive set of data entry or statistical analysis tools. If further analysis or data manipulation is required, a separate database or statistical program, such as Epi Info<sup>1</sup>, must be used. Only specialised procedures not readily available elsewhere have been included.

The use of each of these programs in Survey Toolbox is described in the book. When the use of one of the programs is discussed, a computer icon such as that




---

<sup>1</sup>EpiInfo 6.04 (DOS) or EpiInfo 2000 (Windows 95) are specialised programs for epidemiological analysis, produced by the Centres for Disease Control, Atlanta, GA. These programs are distributed free, and have been included on the CD version of the software accompanying this book. If you don't have the CD, you can obtain a copy through the Internet from <http://www.cdc.gov>. The use of these programs is described briefly in Chapter 6.

shown appears in the left margin, and the program name appears in **Bold Type**. A listing of the name and use of each program, as well as a reference to the pages where it is described, is given below, while a full description of the purpose, data input and outputs of each program is given in Appendix C.

## Computer Programs

Program Name	Purpose	Page
Random Village	Random selection from a list of villages	50
Random Animal	Random selection of animals in a village	58
RGCS Win95	Selection of random geographic points	68
RGCS ArcView	Random points with map interface	69
Prevalence	Sample size and analysis of two-stage prevalence surveys	160
Compare Prevalence	Compare the results of two prevalence surveys	167
True Prevalence	Convert apparent to true prevalence	165
Survival	Survival analysis for disease outbreak surveys	177
Survival Size	Sample size for disease outbreak surveys	173
CapRecap	Two-sample analysis for incidence rate estimates	187
FreeCalc	Sample size and analysis of surveys to demonstrate freedom from disease	198
Survey Toolbox	Main menu giving access to all programs	5



**Survey Toolbox** is the main menu, shown below, and gives you quick and convenient access to all of the programs. It provides the main link between the book and all the programs.

**Requirements**

Windows 95

The programs have been written to run under the Windows 95 and later compatible operating systems (IBM compatible computers). A full installation (including the electronic copy of this text) occupies 5 megabytes of hard disk space. If only the programs are installed, they occupy 2 megabytes.

MS-DOS

Some of the programs are also available in an MS-DOS version for older computers (as indicated in the list in Appendix C). These are located on the CD in the \MSDOS directory.

**Installing**

If you are installing from the CD, or floppy disk, insert the disk, click on the **Start** button, and select **Settings | Control Panel**. From the Control Panel, double click on the **Add/Remove Programs** icon, and click the **Install** button. The Windows Install program will guide you through the steps.

For the MS-DOS version of some of the programs, change to the \MSDOS directory (if using the CD), or insert disk 3 (if using floppy disks). Create a new directory on your hard drive called \ToolBox and copy all files into that directory.

**Running**

If using the Windows 95 version, the install program will create a new entry in your Start menu, called Survey Toolbox. Click on this to open a list of the different programs, then click on the program you want. You may want to place a shortcut to the **Survey Toolbox** menu on your desktop, to make it easier to access (consult Windows Help to find out how to make a shortcut).

**Using the book**

The book is divided into four parts.

Part I, *Background to Disease Surveys* provides detailed information on the various techniques and procedures involved in disease surveillance. National epidemiologists designing surveys should review this information, as should those involved in training survey field staff. District veterinary staff responsible for field work may wish to skip some sections but will find useful information on sampling animals in villages, restraint, specimen collection, and village interviews in Chapters 3, 4 and 5.

Part II, *Guidelines for Conducting Surveys* provides a detailed description of the three main survey types described in this book. This is essential reading for national epidemiologists and those responsible for planning and organising disease surveys. Provincial and state staff responsible for the field implementation of the survey should read the chapter on the relevant survey type, and some sections will also be valuable to district field staff.

Part III, *Notes for Trainers*, is specifically targeted at those wishing to use this book as the basis for training of survey staff. It contains guidance on who should become trainers, and effective teaching techniques. This is followed by a series of lesson plans, divided into three training courses. There is also a collection of activity sheets to help organise participatory training activities.

Part IV, the *Appendices*, contains additional information that may be useful to some readers. This includes a glossary of statistical and epidemiological terms used in the text, a list of statistical formulae used (implemented in the software), a list of the software programs, example data collection forms which can be copied and used, and a list of the contents of the CD or floppy disks.

The book is designed to be flexible to use. For those needing to understand the survey procedures and all necessary background in depth, starting at Chapter 1 and reading through to the end will provide a thorough understanding. Many of the target users will not require all the information, and so might like to pick and choose. There are three ways to do this:

- Use the contents to find the chapter you want, or the index to find the particular reference you want,

- Look at the adjacent flow chart (also shown on the inside front cover). This will guide you to references to other flow charts or instructions throughout the book that help you decide what information you need or how to carry out a survey procedure. The flow charts refer to the page where you can find the information.
- Refer to the table below, which lists the recommended sections for different types of readers.

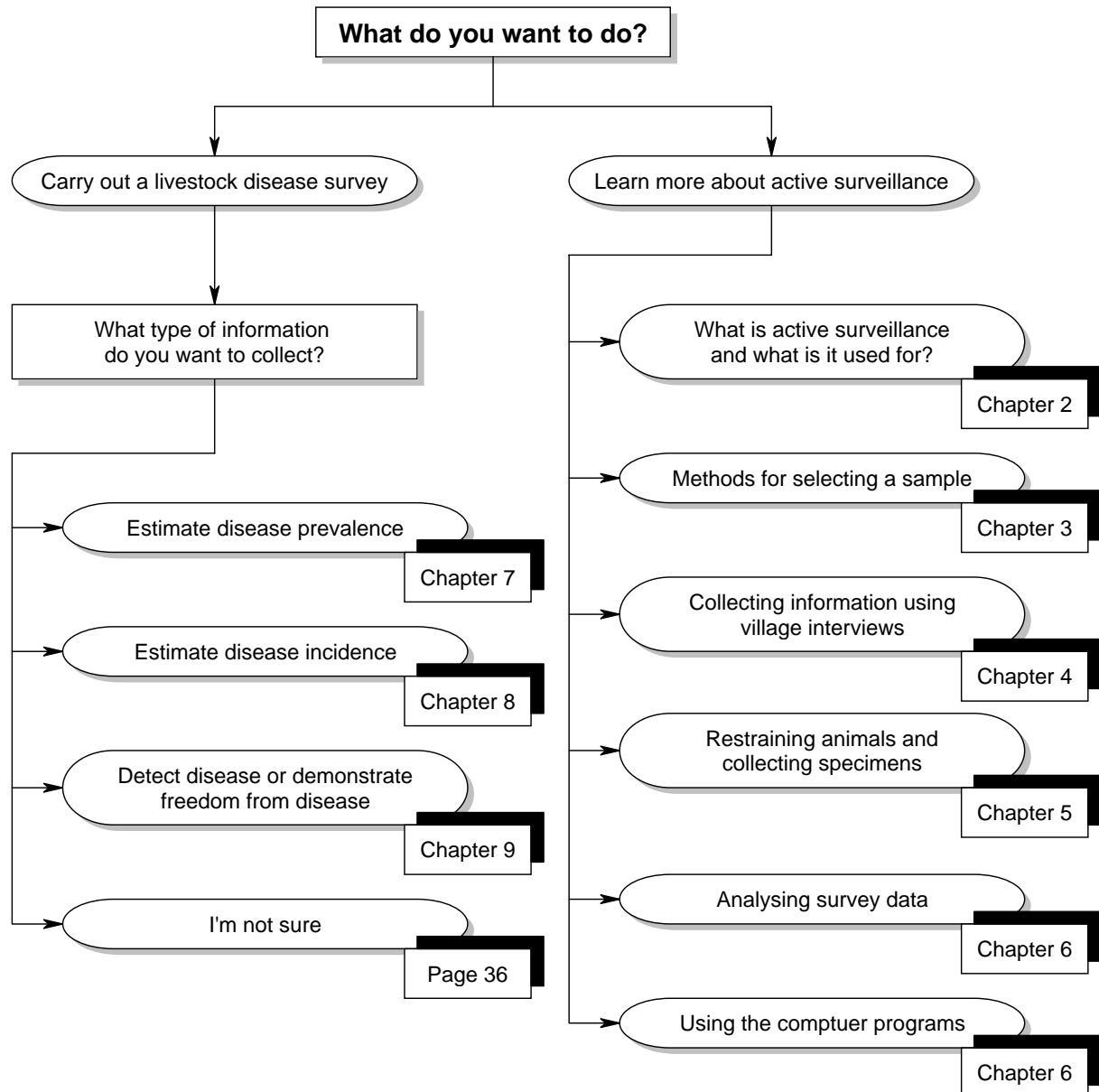
Conducting a large scale survey requires the involvement of a number of field staff. The book is also designed to be used as a teaching resource as part of training courses. Part III is specially designed for trainers, and provides notes, exercises, and suggested lesson plans for training courses. Much of the material throughout the book has been designed to make it easy to copy and use as training material.

8 Chapter 1: Introduction

What chapters should you read?

Chapter / Section	National Epidemiologist or Development Worker	Trainer responsible for training field survey staff	Provincial or State staff coordinating surveys	District staff performing field survey activities
1. Introduction	✓	✓		
2. Principles of Surveillance	✓	✓	✓	
3. Sampling	✓	✓	✓	✓
4. Data Collection	✓	✓	✓	✓
5. Village Interviews	✓	✓	✓	✓
6. Data Management	✓	✓	✓	
7. Prevalence Surveys	✓	✓	✓	✓
8. Incidence Rate Surveys	✓	✓	✓	✓
9. Freedom from Disease	✓	✓	✓	✓
10. Trainers Notes		✓		
11. Lesson Plans		✓		
12. Activity Sheets		✓		

✓ Should read this section  
 ✓ Read this section only if conducting the relevant survey type



## Part 1

### Background to Disease Surveys

The chapters in the first part of the book introduce some of the important ideas to conducting livestock disease surveys. A range of concepts and practical techniques is discussed, covering all aspects of the survey process.

Chapter 2, *General Principles of Animal Disease Surveillance*, introduces some of the principles of surveillance and the need for livestock surveys. Some of the important concepts behind surveys are discussed as well as different measures of disease.

Chapter 3 deals with survey sampling. Selecting a sample is one of the most difficult parts of conducting a survey that will accurate results, so as well as providing background information, a range of practical approaches to sampling is discussed.

When the sample has been selected, the survey, the survey field work needs to be carried out, and the information collected. Chapter 4, *Principles of Data and Specimen Collection*, provides guidelines for collecting interview data, restraining animals, and collecting specimens.

Village interviews play a key role in the techniques described in this book. They can be used to collect good quality information quickly and inexpensively. Chapter 5 is devoted to a discussion of techniques for conducting a successful village interview.

When the data have been collected, they need to be analysed to produce useful information. Computers are an integral part of the techniques described in this book, as they are able to free the survey staff from the need for high-level statistical skills. Chapter 6, *Computerised Data Management*, introduces the basic skills required to manage and analyse data using a computer. Specific analytical techniques for the different survey designs are covered in Part II.

# 2

## General Principles of Animal Disease Surveillance

### Contents

Chapter 1: Introduction

#### Part I: Background to Disease Surveys

Chapter 2: General Principles of Animal Disease Surveillance  
Animal health information  
Disease surveys  
Measures of disease  
Diagnostic tests  
What survey to use

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### Part II: Survey Design and Analysis

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### Part III: Notes for Trainers

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



Users of animal health information

## Animal health information

Information on the health status of livestock is needed by a wide variety of people and organisations, including livestock owners, district, provincial, and national veterinary authorities, businesses, research organisations, and regional and international animal health organisations. Each uses the information for somewhat different purposes. This discussion will focus on the needs of national and sub-national veterinary authorities.

### Why do veterinary authorities need animal health information?

The role of the national veterinary authorities is to control animal diseases, improve the health and productivity of the nation's livestock, and thereby, the well-being of the people. To achieve this, information is required to:

- identify what diseases exist in the country;
- determine the level and location of diseases;
- determine the importance of different diseases;
- set priorities for the use of resources for disease control activities;
- plan, implement and monitor disease control programs;
- respond to disease outbreaks;
- meet reporting requirements of international organisations (e.g. OIE);
- demonstrate disease status to trading partners.

### Collecting animal health information

Passive reporting

#### Passive surveillance

The main method of collecting information on livestock diseases currently used in most countries is through a passive disease reporting system. When an animal is noticed sick, the owner may contact the veterinary authorities, who may then either submit a disease report, or send a specimen to a diagnostic laboratory. These reports and/or the results of examination of the specimens provide information on what diseases are present in the country. Information is not collected about all diseases, or all cases of disease. Many countries have compulsory disease notification regulations to encourage the reporting of priority diseases, but other diseases are not reported.

This system is called a passive reporting system or *passive surveillance* because the main users of the information (the veterinary services) take no action to initiate the collection of the information. The livestock owner initiates the report, and the central veterinary authorities wait (passively) for the report to arrive. Passive surveillance also includes the use of information that was collected for a different purpose, such as diagnosis.

Passive surveillance is a system in which veterinary authorities make no active efforts to collect disease information; they just wait for disease reports to come to them.

Passive reporting systems provide important information needed by the veterinary authorities. Reports, especially those supported by laboratory diagnosis, provide information about which diseases are present in the country, and in what areas. They also provide information needed to respond to outbreaks of disease. Passive surveillance does have some problems, however.

Under-reporting

The most important problem is that of *under-reporting*. Even when regulations make it compulsory to report a particular disease, the report depends on several people: firstly the owner must recognise that the animal is sick, and then notify the livestock officer. The officer must in turn submit a report to the central authorities or diagnostic specimen to the laboratory. In many countries, there are several more steps, where, for instance, the report is passed through provincial or regional offices as well, depending on even more people. The weakest link in the reporting chain is usually the livestock owner, who may not recognise the disease, or may fail to report it for other reasons. The result is that not all cases of disease are reported. It is rarely possible to estimate the level of under-reporting (although Chapter 8 describes one method), so it is impossible to calculate the total number of cases of disease.

**Example:** Two cities, Dogtown, and Canineville, are trying to control canine rabies. Both have about the same number of dogs. The veterinary authorities in Dogtown have decided to start a campaign where free vaccination is provided to all dogs. Canineville has decided that it will continue to charge dog owners for rabies vaccinations. After one year, the records of the veterinary diagnostic laboratories in the two cities are examined. Dogtown, with free vaccination has had 35 confirmed cases of rabies. Canineville has had 78. Has the use of free vaccination been successful?

The lower number of diagnoses in Dogtown suggests that the campaign has been successful. However, there are many other possible reasons why the number of diagnoses is lower. If the proportion of stray dogs in Dogtown is much higher than Canineville, then many stray dogs contracting rabies may die without being noticed or reported. The cost of free vaccination may have meant the city had to reduce the number of veterinary staff. The remaining staff may have been so overworked, that they didn't have time to submit specimens from suspect cases to the laboratory. Alternatively, the laboratories may have been using different tests to diagnose rabies (e.g. Seller's stain in Dogtown, and the Fluorescent Antibody Test in Canineville) resulting in a lower number of positives in Dogtown.

The information from the laboratories cannot therefore be used to conclude whether or not the use of free vaccination has been successful.

Passive reporting systems are often not able to provide information on the total amount of disease, because of under-reporting.

Another problem with passive reporting systems is that there are many different reasons why a report may not be made when an animal gets sick. The reasons can vary between areas, or types of livestock owners.

**Example:** A particular developing country, Mountania, has large areas of mountainous terrain, and a very poorly developed road system. Some commercial cattle farms exist in the fertile plains close to the capital city, but most cattle are raised by smallholder villagers in the mountainous regions. Examination of the disease report records indicates that a large number of reports of foot and mouth disease (FMD) has been received from intensive cattle farms near the capital, but virtually no reports have been received from smallholders in the mountainous regions. What can we conclude about the distribution of FMD in this country?

The records suggest that the level of FMD is much higher in commercial farms than amongst smallholders, and that there is more FMD on the plains than in the mountains. Clearly this conclusion is not necessarily true. Intensive farms near the capital are likely to be in regular contact with the veterinary authorities and report almost every case of FMD. Smallholders in inaccessible parts of the country may not know that they should report the disease, or may be unable to contact veterinary officers to make a report. The patterns of reporting are not reflecting the patterns of disease distribution but differences in the efficiency of the reporting system in different places.

Passive reporting systems usually cannot provide representative information on the level of disease in the population, or the geographical pattern of disease. More reports may come from one part of the population than another.

A third problem with passive reporting systems is that the size of the population that the disease reports relate to is generally not known. This makes impossible the calculation of useful measures of disease, such as rates and proportions.

Rates and proportions allow the comparison of disease figures from populations of different sizes. Two commonly used measures are prevalence (a proportion) and incidence rate, discussed under Measures of Disease (page 26).

**Example:** Two neighbouring provinces, one with a larger number of villages raising cattle than the other, are trying to control haemorrhagic septicaemia (HS). Field reports over the previous year record 36 village outbreaks of HS in the larger province, and 24 in the smaller province. Which province has the highest level of disease?

Although there are more outbreaks in the larger province, there are also more villages with cattle that are susceptible to suffering an outbreak, so you would expect more outbreaks. To compare the provinces, we need to know the total number of villages in each province. The problem is that some of these villages might not submit a report, even if there was an outbreak (perhaps because they are inaccessible, or perhaps because the head of the village has had an argument with the local veterinary officer, and refuses to talk to him). The 36 reports in the larger province have come from villagers which would report an outbreak. The total number of villages which would report an outbreak (but didn't have one) is needed to calculate the rate, but this figure is unknown.

**Rates** are measures of the frequency of an event in a population.  
**Proportions** are measured by percentages.

Passive disease reports are not reliable enough to be used to calculate rates or proportions.

These problems with passive reporting systems limit the value of information collected. However, the information *can* be used to:

Uses of passive reporting systems

- identify which diseases are in the country (but not prove that some disease is not in the country), if diseases are correctly diagnosed;
- identify where the disease is located (but not identify areas where there is no disease);
- respond to disease outbreaks; and
- meet the basic disease reporting requirements of OIE.

On the other hand, passively collected information *cannot* be used to:

Limitations of passive reporting systems

- determine the level and geographic pattern of the disease;
- determine the importance of the disease;
- set priorities for the use of resources for disease control activities;
- plan, implement and monitor disease control programs; or
- demonstrate disease status to trading partners.

### **Active surveillance**

In active surveillance the main users of the information (usually the veterinary authorities) make active efforts to collect the information needed. In contrast to passive surveillance, the main reason for collecting the information is surveillance. As the collection of information is controlled by the users, it is possible to make sure that the information will be of appropriate quality.

Requirements of active surveillance

In order to overcome the problems of passive surveillance, active surveillance must be able to:

- avoid problems of under-reporting;
- collect information which properly represents the true disease situation in the population; and
- comes from a population of a known size to allow the calculation of rates and proportions.

The most practical way to achieve this is through the use of properly structured disease surveys. Surveys have two further advantages: they can be quick to conduct, and relatively inexpensive (compared with the cost of running an effective passive reporting system).

Active surveillance uses structured disease surveys to collect high quality disease information quickly and inexpensively.

## Disease surveys

Census

In order to produce complete reporting (and accurate measures of disease frequency), a passive disease reporting system needs to gather information about every single case of important diseases in the country. To achieve this, every single animal must be regularly examined. This type of data collection is known as a *census*, where every member of the population must be checked. As veterinary services are unable to do this, it is the responsibility of the owners, who have little or no training in disease diagnosis. This is why some owners fail to recognise a disease in the first place.

Survey

Surveys are able to gather reliable information quickly and inexpensively because instead of requiring a census, in which the whole population is examined (by untrained owners), only a small proportion (a sample) of the population is examined (by trained veterinary staff).

A census examines every member of the population. A survey examines only a small part of the population.

## Populations, units of interest and samples

Population

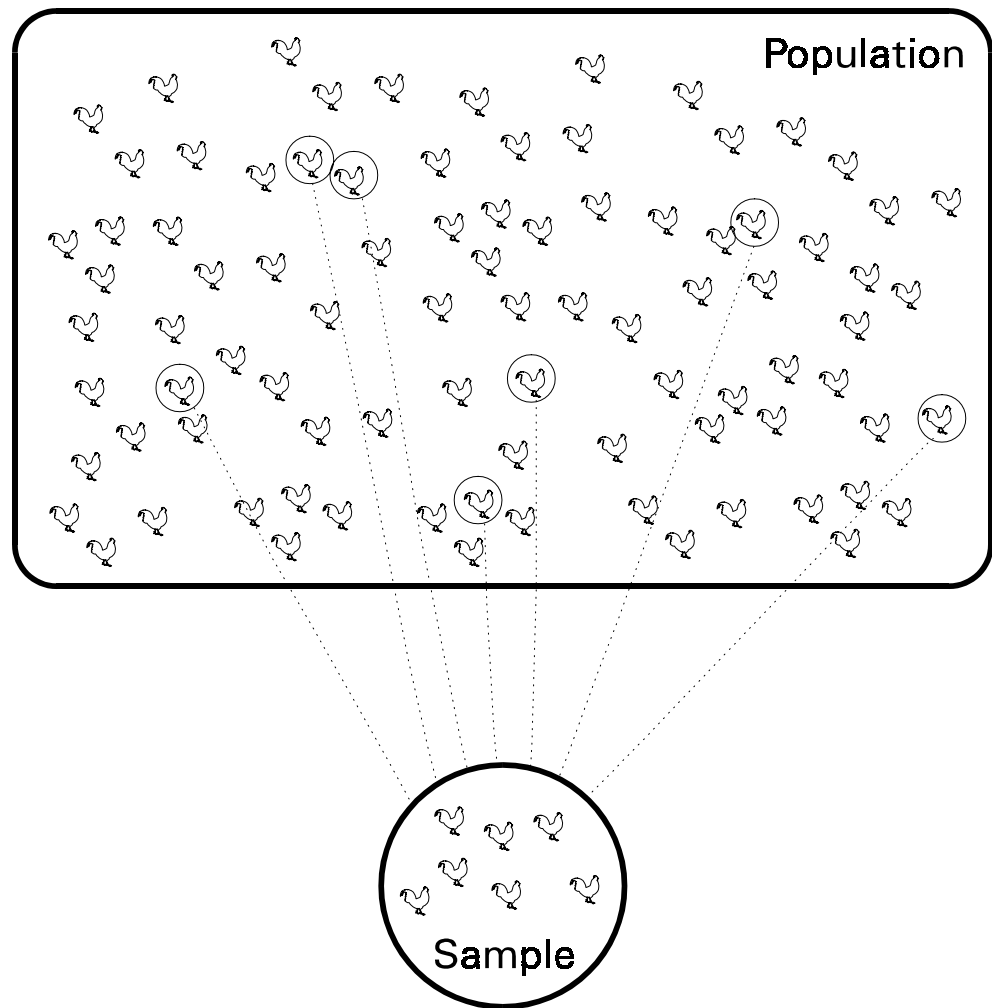
This introduces two important concepts: the population, and the sample. The population consists of all the things in a particular area that we want to know about. The population usually consists of animals, but may consist of other things (villages, farms, livestock owners). The things making up the population are known as the *units of interest*.

Units of interest

**Example:** We wish to know more about goat pox, and decide to conduct a national survey to determine the level of the disease. The objective of the survey is to estimate the prevalence of goats in the country that have goat pox. The population that we are interested in is *all the goats* in the country. The unit of interest is the animal (goat).

**Example:** We are designing a survey to assess the economic impact of mastitis. The objective is to estimate the prevalence of cows with mastitis. To determine the level of disease, we are interested in a population consisting of *all lactating dairy cows*. The unit of interest is the animal (lactating dairy cow).

**Example:** We are studying the occurrence of classical swine fever in village pig herds. We want to calculate the prevalence of villages which have experienced an outbreak of the disease in the last year. The population of interest is *all villages that have pigs*. The unit of interest is the village.



The sample is a small group of units (animals, people, villages) that have been selected from the population. Each element in the sample is then examined to collect disease information.

A survey involves the examination of a small group (a sample) of elements (or units of interest) drawn from all the elements of interest (the population).

There is one problem with surveys. Once we have examined every element in the sample, we know exactly what the disease status of the sample is. However, we know nothing about the rest of the population which we have not examined.

**Example:** A national vaccination campaign has been started to control Aujeszky's disease (Pseudorabies) in pigs. We wish to monitor the effectiveness of the campaign, and conduct a survey to determine the prevalence of pigs with protective antibodies to the disease. There are eight million pigs in the population. Instead of testing all the pigs we take a sample of 200 pigs, collect blood and test the antibody levels. We find

that 164 pigs (82% of the pigs tested) have protective antibodies. What proportion of the population has protective antibodies?

There is no way of knowing what the disease situation in the rest of the population is, as we haven't examined the other 7,999,800 animals. It is possible (although very unlikely) that none of these animals has been vaccinated and that in our sample of 200, we have tested the only 164 vaccinated pigs in the whole country.

## Inference

If surveys tell you a lot about a small number of animals, but nothing about the rest of the population, then what is the value of them? How can we use the results of a survey to learn something about the animals that haven't been examined?

The answer is *inference*. Inference is the process of estimating the true value of the disease status of the population, based on the results observed in the sample. In the above example, we could use inference to assume that the animals that were not tested were the same as ones that were tested. We therefore *assume* that the proportion of the population with protective antibodies is *approximately* equal to 82%. This is the critical difference between the sample and the population. When we have finished the survey, we know exactly what the disease status of the sample is, but we can only estimate disease status of the population.

Inference is the process of assuming that the disease status of the population is similar to the disease status of the sample.

The danger with inference is that your assumptions can be wrong. If we have examined only 200 pigs in the example, then it is very possible that the rest of the pigs are quite different. The true proportion in the population may not be 82% but something completely different.

Representative samples

While inference always runs the risk of being wrong, we can minimise this risk, by ensuring that the sample selected is as similar to the rest of the population as possible. If the sample and the population are essentially the same (with respect to the characteristic of interest, or disease being measured) the sample is said to be a *representative* sample.

**Example:** Let us continue the above example. Most of the country's pigs (seven million) are raised in the smallholder system as village free-range pigs. There are only about one million pigs in intensive piggeries. If all 200 pigs in the sample were drawn from an intensive piggery, they would be unlikely to be similar to the village pigs. The sample would not represent the national population very well. A more representative sample would be one made up mostly of village pigs with a small number of intensive pigs. For instance, a sample consisting of 175 ( $\frac{7}{8}$ ) village pigs and 25 ( $\frac{1}{8}$ ) intensive pigs would be more likely to be representative.

A representative sample is one which is similar to the population. Inference is only valid when a representative sample has been chosen.

Bias

When, on average, the estimate from the sample is different from the true value in the population, the estimate is said to be *biased*. A single estimate from a survey will usually be slightly different from the true value, due to chance. However, if an identical survey was repeated many times, and the average result of the many surveys is different to the true value, the survey technique is said to produce a biased result. Bias can result from many different problems in a survey, most of which can be avoided through careful design.

Systematic error

Bias is caused by systematic error. Systematic error is an error that predictably causes the same type of error for each observation. For example, when weighing chickens, if the scales used are incorrect, and always indicate that the chickens are 300 grams heavier than they really are, then there is a systematic error in the results, and the estimate of the average weight would be biased. This is an example of *measurement bias*. Another important source of bias is *selection bias*, in which the sample selected is not representative, due to selecting animals which are systematically different from the rest of the population.

Bias is the difference between the average estimate from a survey and the real value in the population, caused by systematic error.

Selecting a representative sample is one of the more difficult tasks in any livestock disease survey. One common situation that poses many problems is selecting a representative sample of animals from within a village.

**Example:** Village buffalo are being surveyed to assess their FMD antibody status to monitor the effectiveness of a vaccination program. There are 42 buffalo owners, and a total of 215 buffaloes. Most owners have two buffaloes, but a small number have herds of up to 25 animals. Owners with only two animals mostly live close to the centre of the village, but the larger herds are either on the edge or some distance out of the village. How do you select a representative sample of 10 animals from the population of 215?

There are several methods that are commonly used in this situation:

- **Sample A:** On arriving in the village, either the head of the village or the village livestock officer is contacted. This person then presents the survey team with ten animals.
- **Sample B:** The team goes to the centre of the village, and moves from house to house, collecting blood from every buffalo until they have ten specimens.
- **Sample C:** The team goes to one of the larger herds, and collects blood from 10 of the animals in that herd.
- **Sample D:** The team wanders through the village selecting some animals from larger herds and some from smaller herds.

Each of these approaches is simple and practical, but in each case the sample selected is unlikely to be representative, and the survey results will probably be biased.

In Sample A, the survey team has no idea how the head of the village chose the animals. It is likely that they went to friends and cooperative owners nearby. If some animals were not vaccinated at the last vaccination round, the animals most likely to have been vaccinated are those belonging to friends of the head of the village, and who live near the middle of the village. Livestock owners living outside the village, or who are not friends with the head are much less likely to have been vaccinated, or to be chosen during the survey. The result is that the sample is not the same as the rest of the population, and the results are likely to be biased.

Samples B and C suffer from similar problems. Sample B is not representative because the larger herds on the edge of the village are not represented at all. Sample C has the opposite problem – only one large herd is represented and none of the other cattle. It is possible that this owner doesn't want to vaccinate, while most of the rest of the village does. The survey would indicate that none of the animals in the sample have antibodies, while the true proportion in the village could be 70%.

Sample D sounds better, as both the larger herds on the edge of the village and the smaller herds at the centre would be represented. But there is still a serious danger of bias. This is because, even when trying to be representative, people tend to choose individual animals for different reasons. For example, the survey team may subconsciously select mostly calves and small cows, and avoid the larger bulls. This is because it is much easier to restrain and collect blood from smaller animals. The problem is that young animals are likely to have lower levels of antibodies than older animals, and are therefore not representative of the population. The survey results would be biased, underestimating the true proportion of animals with antibodies.

Random sampling

There is in fact only one way to be confident that the sample chosen is representative of the population. To select a representative sample, it is necessary to ensure that every animal (unit of interest) in the population has the *same* chance of being chosen in the sample, regardless of its owner, location, size or any other characteristic. Sampling techniques of this sort are known as *random sampling*<sup>2</sup>. Random sampling has a number of other important advantages, and is discussed in detail in Chapter 3.

Random sampling means that every element (unit of interest) in the population has the same probability of being selected in the sample. Random sampling is the only way to reliably select a representative sample.

---

<sup>2</sup>To be slightly more precise, simple random sampling, in which each element has the same probability of selection, is just one of a group of *probability sampling* techniques. These share the feature that each element in the population has a known, non-zero (but not necessarily equal) probability of selection. Probability sampling is the only reliable way to avoid selection bias, and is required if estimates of population values are to be valid. Samples chosen with probability sampling may still be unrepresentative due to chance, but on average, they will be similar to the population in all ways.

## Estimation

The aim of surveys is to determine some characteristic of the population (for example, the proportion of animals with antibodies to a disease, or the average number of pigs in villages). As only a sample of the population is examined, and inference used to make assumptions about the rest of the population, it is likely that the value measured from the sample will not be the same as the real value in the entire population. Random sampling is used to minimise this risk. Because we cannot know what the real value in the population is, we use inference to *estimate* it.

**Example:** A village has 4500 chickens. A survey designed to measure what proportion has been vaccinated for Fowl Cholera, is based on a random sample of 20 chickens whose blood is taken and antibody levels measured. Fifteen of the 20 chickens (75%) have positive antibodies to Fowl Cholera. We therefore estimate that the proportion with antibodies in the village is 75%.

### Sample size

When considering the results of a survey like this, it is important to have some idea of how good the estimate is. The number of animals that are selected in the sample (the *sample size*) is one of the most important factors that determines how close our estimate is likely to be to the true population value. In the above example, 20 chickens were chosen. If instead we had chosen 2000 chickens from the 4500 in the village, it is clear that the estimate of the proportion is likely to be much more precise. On the other hand, if we had chosen only four chickens for our survey, we would not have very much confidence that the result was accurate.

Surveys with large sample sizes produce more precise estimates.

If we consider these three possible surveys of the village chickens, with three different sample sizes, 4, 20 and 2000, it is still possible for each of the surveys to produce an estimate of 75%. However, we would be much more confident that the true population value was closer to 75% if we had used a sample size of 2000 rather than 4.

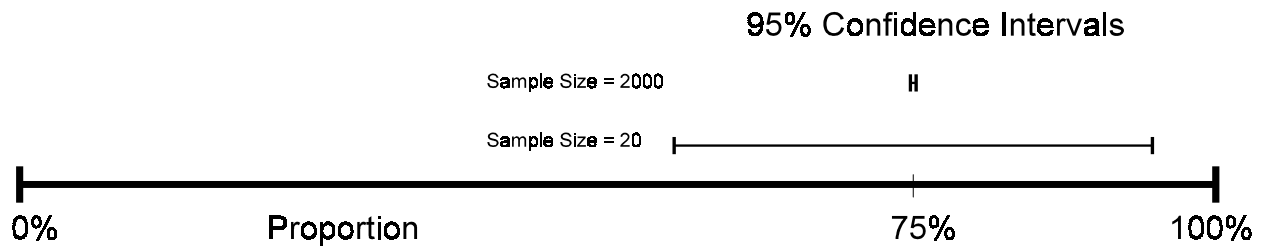
### Confidence interval

When interpreting the results of a survey it is useful to have a measure of how precise the estimate is. This will tell us how much confidence to have in the results. When we use random sampling, it is possible to calculate such a measure, called the *confidence interval*. A confidence interval indicates how close to the real population value our estimate is likely to be. All estimates from surveys should be reported with confidence intervals, so the users know how reliable they are.

The confidence interval of a proportion is a range of values that we are confident that the real value is likely to be in. For instance, in our survey of 20 chickens, with an estimated prevalence of 75%, the 95% confidence interval is 51%-91%. This means that our best guess at the real prevalence is 75%, but we are 95% *confident* that, even if we are wrong, it lies between 51% and 91%. The real value is probably around 75%. It is possible, but much less likely, that it is closer to 51% or 91%. "95% confident" means that if we did the same survey in the same way 100 times, even though we would probably get different estimates each time, the true value would lie in our

confidence interval 95 times out of 100<sup>3</sup>. This confidence interval is quite wide. Although we think that the value is around 75% it could well be as low as 51%, which is quite a big difference.

Consider the same survey, using a sample size of 2000 instead of 20. If we observed 1500 animals with titres, then our estimate would still be a proportion of 75%. However the 95% confidence interval would be 73% - 77%. This means that we are 95% confident that the real value is between 73% and 77%. We can put a lot more faith in the second survey than the first as our estimate is much more precise. This is shown below.



By convention, the precision of an estimate is usually described by the 95% confidence interval. It is possible to calculate confidence intervals at different levels of confidence, such as 90% or 99%, but these are used much less commonly.

A confidence interval indicates how confident we are that the estimate is correct. We can be 95% sure that the true population value lies within a 95% confidence interval. The smaller the confidence interval, the better the survey.

## Survey accuracy

In summary, the accuracy of the estimate from a survey is determined by two factors - *precision* and *bias*.

Precision  
Random error

If the same survey is performed on a population many times, the answer will be slightly different each time. This difference is called random error, and is represented by the width of the confidence interval. If the differences between survey results are small, then there is low random error, and the survey results are very precise. Precision is determined mainly by sample size. A survey with a large sample size will have less random error, and be more precise.

Bias  
Systematic error

If a survey is repeated many times, and the result is always different from the true value by about the same amount in the same direction, this is called systematic error. Systematic error causes biased results, and this is controlled mainly through good survey design.

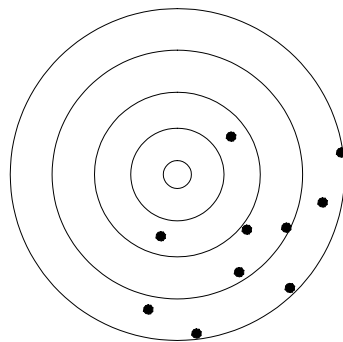
---

<sup>3</sup>The probabilistic interpretation of a 95% confidence interval is as follows. If a survey using the same methodology and sampling strategy was used to study the same population many times, and a confidence interval calculated in the same way based on the result of each survey, the true population parameter would fall within the confidence interval 95% of the time.

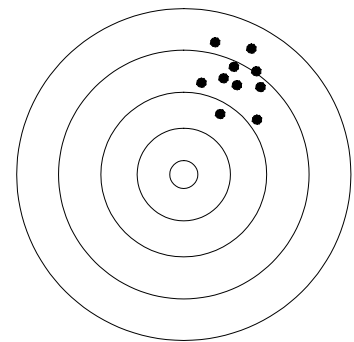
**Example:** Conducting a survey to estimate a value in the population is like shooting a gun to hit a target. If you are not a very good shot, you will hit all over the target. Your shooting is not precise, because there is a lot of random error. In order to overcome the random error, you need to use a lot of shots (use a large sample size) before you are likely to come close to the centre of the target. If you are very good, all the shots will hit very close to the same point. There will be very little random error, and the result will be much more precise. However, you can hit the centre of the target only if the sights on the gun are good. If the sights on the gun are not straight, then when you aim at the centre of the target, you will hit off to one side, and the result will be biased. If there is no bias, then you will hit the centre. This is shown in the figure below.

The real population value that we are trying to estimate is the centre of the target. With each survey (or gun shot) we get a result which may be close to the true value or a long way from it. If the average result of all the surveys (middle of the pattern of gun shots) is close to the true value (centre of the target) then there is no bias, as shown in the bottom two figures. If, however, the average result is centred away from the true result (top figures) bias is present, caused by systematic error (for example, incorrectly set gun sights).

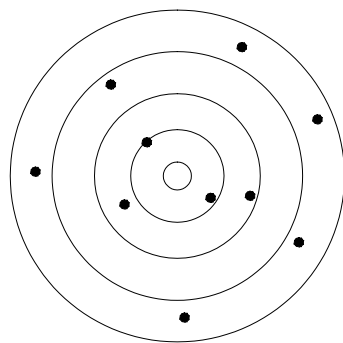
If there is a wide spread of different results from different surveys, it is due to random error, as shown in the two figures on the left. A very good shot will have less random error and a more precise result, as shown on the right.



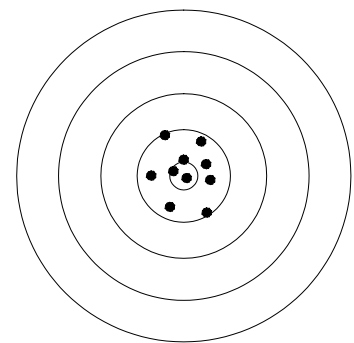
Low precision, biased



High precision, biased



Low precision, unbiased



High precision, unbiased

## Measures of disease

In order to control disease effectively, we must first understand the distribution of the disease: how much disease there is, where it is, in what animals, and so on. Disease surveys are based on counting the number of animals with a disease, and the numbers without. These counts can then be used to calculate several different measures of disease, each describing the level of disease in a different way. The two most important measures of disease for active surveillance are *prevalence* and *incidence rate*.

### Prevalence

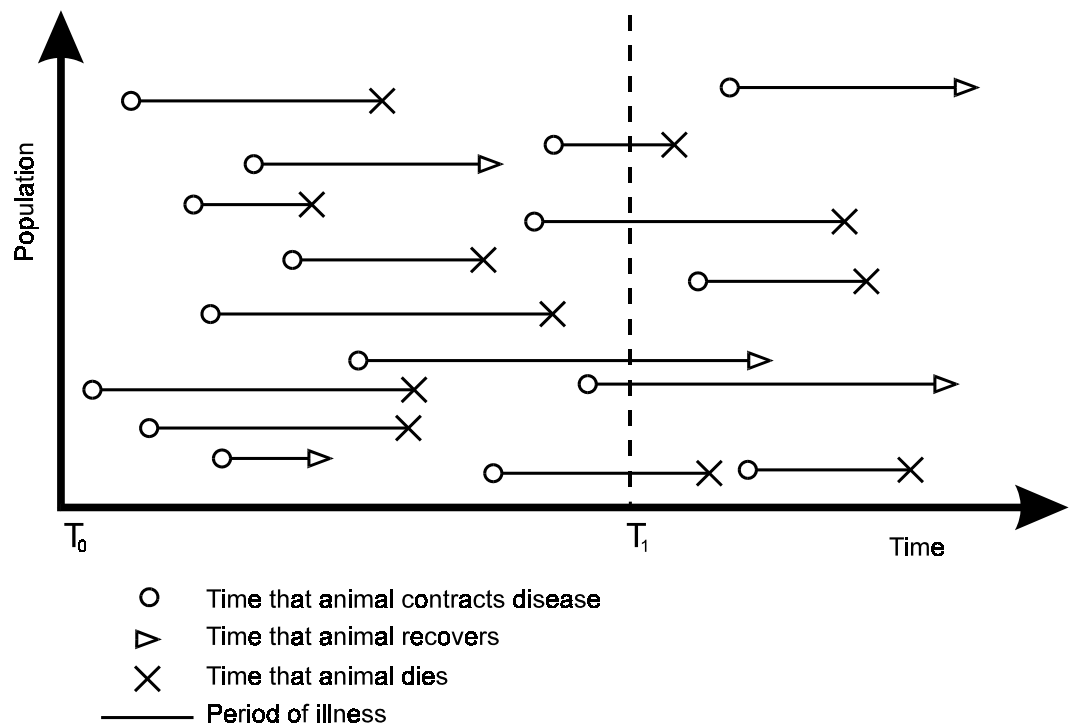
Prevalence (sometimes called point prevalence) is a measure of the number of animals with the disease of interest at a specific time, as a proportion of the total number of animals in the population.

**Example:** A small intensive chicken farm with 2000 birds suffers an outbreak of Newcastle Disease. The first birds start to get sick on the 3 March. By 5 March, many birds are dying. The owner contacts his local veterinary officer who visits on the 6 March. On that day, the officer counts 56 birds showing signs of disease, and the owner reports that a further 143 have already died, and 28 birds had been sick but recovered. There are 1801 apparently healthy birds remaining. What is the prevalence of Newcastle Disease on the farm on 6 March?

The number of animals with the disease is 56. The total number of animals in the population is 1857 (2000 - 143 birds which have already died). The prevalence on 6 March is therefore:

$$\text{Prevalence} = \frac{56}{1857} = 3\%$$

The figure below illustrates the idea of prevalence using a small part of the population over a period of time. Each line represents a single animal with disease. The line starts when the animal first gets sick, and stops when it dies or recovers.



In the figure, prevalence is measured at time  $T_1$ . At that time, the number of sick animals is counted, giving a total of 5. If there were 16 animals in the population at time  $T_0$ , and 6 had died before time  $T_1$ , the remaining population is only 10 animals. The prevalence would be  $5/10$  or 50%. Note that prevalence is measured at a particular point in time, and there is no time period used in its calculation.

Prevalence is the number of sick animals at a single point in time as a proportion of the total population at risk at that time.

$$\text{Prevalence} = \frac{\text{Number of cases at one point in time}}{\text{Population at risk at same point in time}}$$

## Incidence rate

Incidence rate (specifically *true incidence* or *incidence density rate*) is a measure of the average speed at which the disease is spreading<sup>4</sup>. Incidence rate is the total number of *new* cases of disease divided by the total time that each animal in the population was at risk of getting the disease. For simplicity this can usually be calculated as:

$$\text{Incidence} = \frac{\text{Total new cases of disease during a period of time}}{\text{Average number of animals at risk} \times \text{Time period}}$$

In our example of the outbreak of Newcastle disease on the chicken farm, we can use the figures to calculate the incidence rate, or rate of progress of the outbreak. If we use the time period from the beginning of the outbreak (3 March) to the date of the visit (6 March) then the time period is 4 days (inclusive). The total number of new cases of disease during these 4 days is equal to:

- the 143 birds that died,
- plus the 28 birds that got sick and recovered,
- plus the 56 birds that were sick at the time of the visit,

giving a total of 227 new cases of disease. The average number of birds at risk can be calculated by taking the average of the number at the beginning of the time period and the end of the time period. At the beginning (on 3 March), there were 2000 birds. At the end (6 March) 227 of those birds had either died, or had already had the disease, and so were not at risk of getting the disease again. The population at risk was therefore  $2000 - 227 = 1773$ . The average population at risk was  $(2000 + 1773)/2 = 1886.5$ . The incidence rate can therefore be calculated and expressed in several different ways:

$$\begin{aligned} \text{Incidence} &= \frac{227 \text{ new cases of disease}}{1886.5 \text{ at risk} \times 4 \text{ days}} \\ &= 0.03 \text{ cases per chicken per day} \\ &= 21 \text{ cases per 100 chicken - weeks at risk} \\ &= 21 \text{ cases per 100 chickens per week} \end{aligned}$$

What do these numbers mean? The first number, 0.03 cases per chicken per day means that if we look at one chicken for one day, on average we will get 0.03 cases of disease. This is obviously meaningless when we talk about a single animal. The value for incidence rate can be multiplied by a larger number of animals or a longer time to make it easier to understand. The second figure, 21 cases per 100 chicken-week, means that if we have 100 chickens, we could expect 21 of them to get sick in

---

<sup>4</sup>Another commonly used measure of incidence is *cumulative incidence*, which is equal to the number of new cases of disease as a proportion of the total number at risk at the beginning of an observation period. At the individual animal level, it estimates the risk of getting the disease. Both incidence density rate and cumulative incidence depend on a measure of the total number of cases of disease in a particular time period, as discussed in chapter 8. The text deals only with the incidence density rate, but the cumulative incidence can be calculated if desired.

the space of a week, if the rate of spread of the disease stayed the same as it was during the first 4 days of the outbreak.

We can use the diagram above to see how incidence rate differs from prevalence. To calculate the incidence rate, we need to count all the *new* cases of disease in a given time period. The number of new cases of disease between time  $T_0$  and time  $T_1$  is 13. Some of these cases died, some recovered, and some were still sick at  $T_1$ , but we are interested only in the number of new cases, not what happens to them. If there were 26 animals in the population at time  $T_0$  and 20 left at  $T_1$  (two of which had already been infected and recovered, so were not at risk), then the average number at risk over the period was  $(26 + (20-2))/2=22$ . If the time period between  $T_0$  and  $T_1$  was 1 month, then the incidence rate is:

$$\begin{aligned} \text{Incidence} &= \frac{13 \text{ new cases of disease}}{22 \text{ at risk} \times 1 \text{ month}} \\ &= 0.59 \text{ cases per animal - month} \\ &= 59 \text{ cases per 100 animals per month} \end{aligned}$$

Incidence rate measures the number of new cases over a period of time

## Prevalence versus incidence rate

Relationship between  
incidence rate and prevalence

Many other measures of disease exist, but prevalence and incidence rate are the most useful. The two are related, according to the duration of disease. A disease with a high incidence rate but of very short duration will have a relatively low prevalence. A disease of relatively low incidence rate with a long duration will have a high prevalence.

**Example:** A study is conducted of dairy cows in a high producing, year round calving area, to examine the occurrence of subclinical low blood calcium around calving time. The study population consists of all mature dairy cows in the area. Over a period of one year, it is found that the incidence rate of the disease is very high, with 85 cases per 100 animals per year. However, the duration is usually very short, lasting only a day or so. At any one point in time, the prevalence (proportion of cows with low blood calcium at that time) is very very low, at only 0.3%.

**Example:** We are interested in assessing the effects of liver fluke in buffalo in an irrigated rice-growing area. Most animals are infected when they are young, and keep the infection at low levels for the rest of their lives. The incidence rate (number of *new* infections) is relatively low at 8 new cases per 100 animals per year, as most animals already have the infection, and only young calves are susceptible to new infections. Because the duration of the disease is virtually lifelong, the prevalence is very high, at 97%.

The reason prevalence and incidence rate can be so different is because they are describing different aspects of the disease. If the size of the population is not

changing, and the level of disease stays about the same, then we can estimate the prevalence, if we know the incidence rate and the average duration of the disease.

Under certain specific conditions when prevalence is low (< 10%)  
 $\text{Prevalence} \approx \text{Incidence rate} \times \text{Duration of disease}$

**Example:** Lameness in cattle caused by foot-rot has become a problem in a village during unusually wet conditions. The incidence rate of the disease is 5 cases per 100 cattle per month. Infected cattle usually recover in about 1 month (the duration of the disease). The prevalence of the disease is therefore about 5 cases per 100 cattle per month times 1 month, or 5%.

Features of incidence rate and prevalence surveys

When planning a survey, we must decide which measure is most useful for a particular purpose. This is based both on the type of information that is collected and practical considerations.

**Example:** A large national vaccination program for cattle, buffalo and pigs is being used in a country, Vaccineland, to control foot and mouth disease (FMD). The aim is to vaccinate every animal twice yearly. The veterinary authorities wish to monitor the program to see if it is achieving its objectives. They need to decide whether to use prevalence or incidence rate as a measure.

In this example, the aim of the program is to ensure that all animals in the country are protected against infection with FMD. This is achieved through vaccination to provide protective antibodies. The veterinary authorities decide that the best measure to assess the program is the prevalence of animals with protective antibodies to FMD. Instead of a measure of the proportion of animals vaccinated, they decide they want to know the proportion that were *successfully* vaccinated. This allows them to evaluate the combined effects of a range of problems with vaccination program (for example, low vaccination coverage or poor vaccine handling resulting in the use of expired or inactive vaccine). If the prevalence is very high, it means that most animals are protected, and there will be very few outbreaks of FMD. If the prevalence of protective antibodies is low, there are many animals that are susceptible to infection, and the program is not working properly. They therefore conduct a national survey of cattle, pigs and buffalo, collecting blood and testing the antibody levels, to determine the prevalence of animals with protective antibody titres to FMD.

**Example:** Continuing the previous example, the survey is conducted and the results analysed. Nationally, 62% of animals have protective antibody titres to FMD. The authorities conclude that the program is not yet completely effective, but is making significant progress, and decide to continue the program as it is. In order to achieve effective control, a regional organisation has been established to help communication and coordination of control efforts between neighbouring countries. As part of this process, a neighbouring country, Diseaseland, conducts a similar survey, and reports the results to the Vaccineland authorities. Diseaseland

suffers from frequent outbreaks of FMD, but is not in a position to start a vaccination program, so virtually no vaccine is used in the country. The results of the Diseaseland survey indicate that 64% of the animals have protective antibodies to FMD.

Antibodies from vaccination  
and antibodies from natural  
infection

The problem with this example of a seroprevalence survey is that the antibody test for FMD is unable to distinguish between antibodies that have come from vaccination and antibodies from natural infection. In Diseaseland there is no vaccination, so all the animals with protective antibodies (64%) have antibodies because they have been recently infected with FMD. The animals with antibodies in Vaccineland may have got the antibodies from vaccination, but could have got them from being infected as well. Although the measure of antibody prevalence is useful, it doesn't tell the Vaccineland authorities if their vaccination program is causing the antibodies, or if there are still widespread FMD outbreaks causing the antibodies. To answer this question, they decide to do a disease incidence rate survey.

**Example:** The Vaccineland authorities conduct a national incidence rate survey to measure the number of village outbreaks of FMD. At the same time, they ask the veterinary authorities in Diseaseland to run a similar survey. The Vaccineland survey finds that the incidence rate of village FMD outbreaks is 3.4 per 100 villages per year. In Diseaseland they find that the incidence rate is 48 per 100 villages per year.

Based on these results, the authorities in Vaccineland can conclude that most of the antibodies observed in the prevalence survey are due to vaccination, rather than natural infection, and that the program does seem to be working.

In many situations, prevalence surveys are adequate to understand the level of disease in a population, sometimes incidence rate surveys may need to be used as well or instead. This example of FMD demonstrates how both survey types can be used together to help understand the problem when laboratory tests are unable to distinguish between naturally acquired and vaccine antibodies.

The following table compares the two types of surveys:

	Prevalence Survey	Incidence Rate Survey
Cost	High	Low
Speed	Slow	Fast
Unit of Interest	Animal	Village /Herd
Data Quality (specificity)	Good	Moderate
Identifies	"Disease"	History of Outbreaks

## Seroprevalence versus clinical prevalence

The above series of examples illustrates the use of a seroprevalence survey and the relationship between prevalence and duration of disease. Prevalence surveys can examine any aspect of disease – clinical signs of disease, evidence of subclinical

disease, or evidence that the animal has previously had the disease, but has now recovered. In each case, the duration of disease (or evidence of disease) is different. Clinical disease often lasts a relatively short time. Subclinical disease may precede clinical disease for a longer period. Evidence of previous exposure to a disease, in the form of antibodies, lasts for a much longer time after the disease.

**Example:** The duration of clinical FMD is only a few weeks. Subclinical infection can be detected for a day or two before clinical signs, and in some cases in carriers, for several months after infection. The duration of antibody evidence of disease after infection lasts for several years. This means that, for a fixed incidence rate of FMD, the prevalence of clinical disease will be low, subclinical disease higher, and the prevalence of antibody evidence of previous exposure will be highest.

When conducting prevalence surveys, to calculate the prevalence of a disease that is rare (has a low prevalence), a large number of animals must be examined before a significant number with the disease are found. For diseases of high prevalence, fewer animals need to be examined, making the survey faster and less expensive.

This is one of the reasons why *seroprevalence* surveys (surveys to estimate the prevalence of animals with high levels of antibodies indicating previous exposure to the disease or previous vaccination) are commonly used in livestock disease control programs.

It is usually easier (requires a smaller sample size) to measure the level of disease using a seroprevalence survey than a clinical prevalence survey.

One advantage of clinical prevalence surveys is that only clinical examination of animals is required, with no need to collect samples for laboratory examination. This may be faster and less expensive, but laboratory tests are usually able to assist in making a more reliable diagnosis. Other surveys may measure the prevalence of animals shedding specific parasite eggs in their faeces, or with detectable antigens from a particular organism as detected in laboratory testing.

## Diagnostic tests

Many livestock disease surveys require the use of laboratory diagnostic tests to examine specimens collected from the animal. One example is the use of laboratory tests such as ELISAs (Enzyme-Linked ImmunoSorbent Assays), Complement Fixation Tests (CFTs), or Agar Gel Immunodiffusion Tests (AGIDs) to identify the presence of antibodies or antigen in blood.

Very few laboratory tests are perfect, although most tests give incorrect results only occasionally. When using laboratory tests as part of a disease survey, it is important to understand how accurate the test is, and what errors are likely to occur.

## Sensitivity and specificity

The performance of a test is described by two measures – the *sensitivity* and the *specificity*. The sensitivity of a test measures the proportion of truly diseased animals that the test correctly identifies as diseased. The specificity measures the proportion of non-diseased animals that the test correctly identifies as non-diseased.

**Example:** A new diagnostic test to detect antibodies to classical swine fever is being evaluated, and is used to test 20 animals. Ten of the animals are already known to have had the disease and should have high levels of antibodies. The other ten animals come from a disease free area and are known to have no antibodies. When testing is finished, 8 of the 10 antibody positive animals return positive results, and 2 are negative. Seven of the antibody negative animals return negative results, and 3 are positive. The results are summarised in the table below.

Calculating sensitivity and specificity

		True "Disease" Status (Presence of antibodies)		Total
		Positive	Negative	
Antibody Test Result	Positive	8 True positive	3 False positive	11 test positive
	Negative	2 False negative	7 True negative	9 test negative
Total		10 disease positive	10 disease negative	20

Using these figures, we can calculate the sensitivity and specificity of the test.

$$\text{Sensitivity} = \frac{8 \text{ true positive test results}}{10 \text{ disease positive animals}} = 80\%$$

$$\text{Specificity} = \frac{7 \text{ true negative test results}}{10 \text{ disease negative animals}} = 70\%$$

Test sensitivity

The sensitivity of a test is the proportion of truly diseased animals in the population which is correctly identified as diseased by the test.

Test specificity

The specificity of a test is the proportion of truly non-diseased animals in the population which is correctly identified as non-diseased by the test.

**Example:** The sensitivity and specificity of individual tests may vary somewhat depending population being tested and the laboratory doing the testing. General estimates should therefore be used with caution. If the test is being used in an area for the first time, it is better to evaluate it in the local population first. As an example of test characteristic estimates, the brucellosis Rose Bengal Test (RBT) has an estimated sensitivity of about 97%, and a specificity of about 93%.

## Apparent versus true prevalence

If the sensitivity and specificity of a test are known, it is possible to correct for errors that occur because of imperfect tests. When a seroprevalence survey has been conducted, and the specimens analysed, the observed seroprevalence (*apparent prevalence*) may not be correct. This is because some of the diseased animals have been incorrectly identified as being negative, and some of the non-diseased animals have been identified as being positive. These errors can be corrected to give the *true prevalence* with a formula using test sensitivity and specificity.

The **True Prevalence** program performs this calculation for you. By entering the apparent prevalence, sensitivity and specificity, it will calculate the true prevalence. If you also enter the sample size of the survey, it will give you a 95% confidence interval. The program and example calculation are shown on page 165.



## What survey to use?

This chapter has introduced the use of surveys for active surveillance to collect information on livestock diseases, and discussed two important measures of disease. A third survey type, surveys to demonstrate freedom from disease, will be introduced in Chapter 9. The veterinary authorities and survey planners need to decide which type of survey or other data collection system is most appropriate for their needs.

To make this decision, you first need to decide what type of information is needed and what it will be used for. In some cases, this leads directly to the type of survey required. In others, it may be clear that the required information is already available from other sources and no survey is required. Different surveys not discussed in this book may be needed in some situations.

**Example:** You wish to export pigs and are involved in trade negotiations with another country. They are refusing to allow the import of live pigs on the basis that they believe your country has Aujeszky's disease. Although there have been no clinical reports of Aujeszky's disease in the past 5 years, you need to be able to demonstrate that the disease has been eradicated. This situation clearly calls for a survey to demonstrate freedom from disease, described in Chapter 9.

In other cases, the choice of survey type is not so clear. Several different types of surveys may be able to collect relevant information, and in fact more than one type of information may be necessary. For some particular problems, passive data may be the most appropriate source of information.

**Example:** There are moves amongst neighbouring countries to establish a regional FMD eradication program, and you need to decide if the government will participate in this program and fund the control measures. To properly answer the question, it is necessary to know, amongst other things; how much FMD is present; where it is; what the state of current control measures is; what the current impact of the disease is; what the cost of the control program would be; if eradication is feasible and, if so, how long it would take, and what benefits it would bring.

Both prevalence and incidence rate surveys would be required to answer these questions, as well as economic studies (cost-benefit analysis) and other special purpose studies. Passive reporting data may also be useful in making the assessment.

The flow chart on the next page lists some of the possible uses of livestock disease information, and suggests some ways in which this information may be collected. It is designed to act as a guide only, as the final choice of which survey to use will involve many other factors.

# 3

## Sampling

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

The need for random sampling  
Random sampling techniques  
Sampling frames  
Sampling from a sampling frame  
Sampling animals within a village  
Sampling without a sampling frame

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



Sampling

When we wish to collect information from a population, it is rarely possible to test or examine every member of the population (i.e. conduct a *census* - page 18). Instead, a smaller group (a *sample*) is selected from the population, and the members of this group are tested (i.e. a survey - page 18). *Sampling* is the process of selecting this group from the population. Members of the sample are examined, and the results are used to estimate some characteristic of the population from which the sample was drawn.

## The need for random sampling

There are several different sampling techniques, which can be divided into two groups: *probability sampling* and *non-probability sampling*. Probability sampling (random sampling) is the only way to ensure that the sample is representative of the population.

### Non-probability sampling

Problems with non-probability sampling

In non-probability sampling, the probability of a member of the population being selected in the sample is not known, and some groups are more likely to be selected than others. This means that a sample selected using non-probability sampling is unlikely to be representative of the population, and the results of the survey will be biased. You should avoid using non-probability sampling techniques.

Convenience sampling

*Convenience* sampling is an example of a non-probability sampling technique. In convenience sampling, samples are selected because they are easy, quick or inexpensive to collect.

**Example:** An intensive dairy herd of 150 milking cows is surveyed to assess the prevalence of lameness. A sample size of 10 animals is to be examined for lameness. Convenience sampling is used to select the animals: the first 10 cows to enter the shed are selected, because they are easy to examine, and you don't have to wait for all animals to arrive.

The problem with convenience sampling is that the sample rarely represents the population. In this case, lame cows are slower, and are therefore likely to be the last ones to enter the milking shed. A convenience sample of the first 10 cows is likely to have no lame cows, even if the prevalence of lameness is high.

Purposive sampling

Another non-probability sampling technique is *purposive* sampling. In purposive sampling, elements in the sample are selected for some purpose. An attempt may be made to select animals which are judged to be typical of the group. Alternatively, when studying a disease, sick animals may be selected more than healthy ones. Even when 'typical' animals are selected, these are unlikely to represent the range of different animals in the population. Purposive sampling does not produce a representative sample.

Haphazard sampling

*Haphazard* sampling is a technique where elements are selected for no particular reason at all. It is designed to imitate random sampling. Unfortunately, when people select animals, there is always some unconscious reason for each animal being selected. For instance, the person selecting may think "I chose a big one last time, so I will pick a small one this time." The result is similar to a purposive sample, and despite the best efforts, is rarely representative of the population.

Non-probability sampling techniques are unable to reliably select a representative sample. The results from surveys using non-probability sampling are likely to be biased.

### Probability sampling

The term *probability sampling* covers a group of techniques which includes:

- simple random sampling;
- probability proportional to size sampling; and
- random systematic sampling.

#### Simple random sampling

In *simple random sampling* (sometimes just called random sampling), every member of the population has the same chance of being selected. In the example of the lameness survey, if simple random sampling was used, both lame animals and healthy animals, animals that are first in the milking shed and animals that come last would have the same chance of being selected in the sample.

This is the first reason for using random sampling. Samples selected using random sampling are much more likely to be representative of the population than samples selected by non-probability techniques. This means that random sampling can avoid the problem of selection bias, and that estimates of population values made from the sample are more likely to be correct.

On average, random sampling produces representative samples.

The second reason to use random sampling is so we can calculate how reliable our survey results are. When survey results are used to estimate the true value in the population (e.g. the prevalence of lame animals), a formula is used to make the calculation. Similarly, a formula is used to calculate the confidence interval for the estimate, which indicates how confident we are that the results are correct. Each of these formulae is based on the assumption that the sample was selected using random sampling. If the sample was selected with a non-probability sampling technique, the formulae are no longer valid and the results may be incorrect.

**Example:** A prevalence survey was carried out to estimate the prevalence of pigs with antibodies to hog cholera in a village. A sample of forty (40) village pigs (the unit of interest) were selected from a village with a total population of 120 pigs. The survey team asked the chief of the village to take them to 4 different pig owners with more than 10 pigs each, and took blood samples from 10 pigs in each herd. Of the 40 pigs, 12 had positive antibodies to hog cholera. Can you infer what is the prevalence of pigs with antibodies to hog cholera in the village?

The sampling technique used was a non-probability sampling technique (convenience sampling). We know that 30% of the sample had antibodies ( $12/40 = 30\%$ ). We can estimate that the prevalence of pigs with antibodies to hog cholera in the population is also 30%. However, because random sampling was not used, there

is no way to determine how likely that estimate is to be correct. We think the true value is 30%, but it could just as easily be 5% or 80%.

If random sampling had been used, we could calculate exactly how likely the estimate is to be correct. The 95% confidence interval is 17% - 47%, which means that we are 95% sure that the true value lies between 17 and 47.

When random sampling is used, we can calculate how reliable an estimate is.

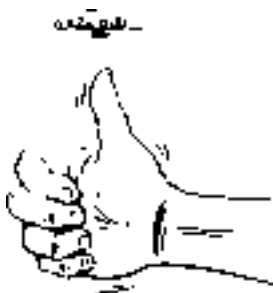
## Random sampling techniques

Random numbers



Random sampling is based on the concept of randomness, and the use of random numbers. *Random numbers* are best explained by an example. Dice have six numbered sides, 1 to 6, and when one is rolled, each side has the same chance of ending on top. However, at each roll, we never know which number will come up. What we do know, is that if we roll the dice again and again, on average, all numbers will appear equally often. Rolling dice is one way of generating random numbers (in this case, between 1 and 6).

Playing cards are another example of randomness. When we shuffle cards, we never know what order they are in. But it is possible to predict what will happen over a large number of games of cards. This is because each card in the pack has the same probability of being on top. That is how casinos are able to make money. They don't know if a particular person on a particular day is going to win or lose, but they do know that most of the time, more people will lose money than win money, because they can predict the average result of a large number of games.



When selecting a sample for a survey, we want to select members of the population in such a way that will ensure that each member has exactly the same chance of being selected. There are many different ways of doing this, and the aim of this chapter is to explain some techniques which are useful to people conducting livestock surveys in developing countries.

### Physical randomisation

The examples of dice and cards are called *physical randomisation* techniques because we actually take physical objects and mix, shake, roll or shuffle them. This represents one of the simplest approaches to selecting a random sample. The problem with dice is that there are only 6 numbers (although decimal dice exist, with 10 sides numbered 0 to 9). Blank cards are much more flexible, as shown by this example.

Blank cards for random sampling

**Example:** A small goat herd is surveyed to determine if it is infected with brucellosis. There are 30 goats (the unit of interest) in the herd (the population), but only 8 are needed for testing (determining the sample size is explained in Part II). Each goat has an ear tag with a unique identification number on it. To select the sample of goats, 30 blank cards

or pieces of paper are used. The ear tag number of each goat is written on a card. The 30 cards are then shuffled well, and 8 cards selected. The goats with the ear tag numbers selected are the ones to be included in the sample.

Problem with physical  
randomisation

This is an effective method for random selection. However, when the group is too large, it can quickly become impractical. For instance, imagine conducting a national survey of village chickens in which chickens from 100 different villages were to be examined. If there are 24,200 villages in the country, this would require writing the village name on 24,200 cards, shuffling and dealing out 100. Shuffling 24,200 cards could be difficult. For this reason, physical randomisation techniques are only used in small surveys. In larger surveys, random numbers are more convenient.

## Random numbers

*Random numbers* are numbers which have been generated randomly, or by chance alone. That means that for each digit, the chance of it being any number between 0 and 9 is the same. There are two sources of random numbers: *random number tables*, and *computer generated random numbers*<sup>5</sup>.

**Example:** A computer was used to generate a random number, 39024. The number has 5 digits, 3, 9, 0, 2 and 4. When the computer selected the first digit, 3, the chance of it being any number between 0 and 9 was exactly the same. The number 3 was just selected by chance. For the second digit, 9, again, the chance of any number between 0 and 9 being selected was the same. The fact that 3 had been selected for the first number made no difference to which would be selected for the second number and so on. It was simply a matter of chance. It is as if the computer is rolling special decimal dice (with 10 sides) and recording each digit as it is rolled.

An example of a random number table is included in Appendix D, and its use is explained below on page 42. Various computer programs, including Epi Info, include random number generators, as explained on page 43. Both can be used equally well to select a sample using the following steps:

Selecting a random sample

**Step 8:** Make a list of all the members of the population. This list is called the *sampling frame*, and is discussed on page 49.

**Example:** A survey is planned of village livestock disease problems in one district. The objective is to identify which disease problem occurs most commonly. There are 75 villages in the district (the population), and a sample of 10 villages is required. The unit of interest is the village. All the villages are listed by name (only the first five are shown).

---

<sup>5</sup>Computer generated random numbers are really “pseudo-random numbers”. They are a sequence generated by a formula, so that if you know the formula, you can predict what the next number will be. With real random numbers, this is not possible. However, for survey purposes, computer-generated pseudo-random numbers are just as good as true random numbers.

	Village Name
	Nong Bone
	Sobtui
	Hang Chat
	Khounta
	Si Meuang

**Step 9:** Number each member on the list from 1 up to N, the total number in the population.

**Example:** The villages are numbered from 1 to 75.

N°	Village Name
1	Nong Bone
2	Sobtui
3	Hang Chat
4	Khounta
5	Si Meuang

**Step 10:** Using a computer or a random number table, select random numbers between 1 and N. Select one random number for each element to be selected in the sample.

**Example:** 10 random numbers are selected, between 1 and 75. The numbers are: 68, 2, 52, 5, 27, 57, 42, 66, 47, and 53.

**Step 11:** For each random number selected, find the corresponding element on the list. These are the ones to be included in the sample.

**Example:** Find the selected villages corresponding to the random numbers.

N°	Village Name	
1	Nong Bone	
2 *	Sobtui	Selected
3	Hang Chat	
4	Khounta	
5 *	Si Meuang	Selected

**Selecting random numbers using a random number table**

Random number tables are a convenient source of random numbers. An example of a random number table is shown below and a full table is given in Appendix D. There are sets of numbers grouped into 5. In the above example, 10 numbers were selected to pick 10 villages from a total of 75. The way to use a random number table to select random numbers is as follows:

Using a random number table

**Step 1:** Choose a starting point and direction. You can start at the top of the table, or you can start anywhere in the middle. You can go across a

row, or down a column. In this example, we will start at the top left number, and move across.

- Step 2:** Calculate the range for your random numbers. The numbers required in this example are between 1 and 75.
- Step 3:** Determine which digits to use from the numbers. The maximum number we want is 75, which has two digits. We therefore only need two of the five digits in each random number. To use the numbers efficiently, we can “cut” them in half, and think of the first two digits (42) as the first number, and the third and fourth digits (53) as the second number. The last digit can be ignored.
- Step 4:** Search through the table for numbers in the required range. Any number between 1 and 75 is counted as one of our random numbers. Any number over 75 is ignored. Continue searching until enough numbers have been found (ten in this example).

**Example:** Using the table below, the first number is 42. This is between 1 and 75, so it is accepted. The second number is 53, and is also accepted. The next digit (9) is ignored. Moving to the right to the next group, the next number is 77. This is greater than 75 so is ignored. The next number is 68, which is accepted as our third random number. The last digit (6) is ignored. Continuing in this way we get a fourth (66), a fifth (52), a sixth (27), discard the next (79), a seventh (02, or 2), and eighth (47), a ninth (57) and a tenth (05, or 5).

Random number table

42539	77686	66524	27792	02474	57058	61530	76108	49436
27030	88085	84744	32591	57804	54790	24545	73422	23337
50253	66592	66151	18506	04391	35824	35397	32031	67780
54127	25147	79021	54189	43708	08178	82187	72106	53795

When using a random number table, it is a good idea to circle the numbers you select, and to cross off those that you discard. This helps you remember which numbers you select, and prevents you from using the same numbers again. You should always use new random numbers when sampling in a survey.

42539 77686 66524 27792 02474 57058

### Selecting random numbers using a computer

While using a random number table is quick and simple, the job can be done even more conveniently using a computer. Various programs are available to select random numbers, and one is included in **Epi Info**. To generate random numbers using Epi Info do the following:

- Step 1:** Start Epi Info, and select EpiTable from the the Programs Menu.
- Step 2:** Open the Sample menu, and select Random Number List.
- Step 3:** Enter the number of random numbers you want (in our example, 10).
- Step 4:** Enter the minimum value for the numbers (usually 1).
- Step 5:** Enter the maximum value for the numbers. This is equal to the total number of elements in the population (in our example, 75).
- Step 6:** Set the program to select either with or without replacement (see ‘Replacement’ below).

**Step 7:** Select the **Calculate** button. EpiInfo will generate random numbers in the required range, and display them on the screen. They can then be printed or saved to a file.



To make sampling even easier, there are specialised computer programs that help you build the sampling frame, select the random numbers, and report the selected elements for you. Two such programs are included with the **Survey Toolbox**. The first, **Random Village**, is for selecting elements from a list, such as villages or herds. The second, **Random Animal** is for a more difficult situation – selecting animals from within a village. Both are described later in this chapter.

### Replacement

Sampling with replacement

Sampling can be done in two ways, either with or without replacement. Sampling with replacement is best illustrated with an example.

**Example:** A standard deck of 52 cards is shuffled, and one card is drawn from the deck. The value of the card is recorded, and then the card is replaced back into the deck. The deck, still containing 52 cards, is shuffled again, and another card drawn. The value of this card is recorded, and it too is replaced. This is continued until enough cards have been drawn.

Sampling without replacement

In contrast to this example, sampling without replacement means that as each card is drawn, it is kept out of the deck, and only the remaining cards are shuffled. The difference between the two techniques is that it is possible to select the same element twice when using sampling with replacement, but it is not possible when sampling without replacement. In some survey designs, it is better to sample with replacement, and in others, it is better to sample without. When the population is large, the difference between the two techniques is unimportant.

If sampling without replacement, using a physical randomisation technique (e.g. dice) or a random number table, you must be careful to check if a number has already been selected. If so, it is discarded, and a new number drawn.

When using **Survey Toolbox** to generate random numbers or do the sampling, you can simply instruct the computer which method to use, depending on the survey design. The survey designs described in Chapters 7, 8 and 9 tell you whether to use replacement or not.

## Systematic sampling

Systematic sampling is an alternative technique when the elements of a population are difficult to individually identify and list, but they can be ordered in some way.

**Example:** A survey is being conducted in a flock of 600 sheep to examine the level of intestinal parasites. If we wish to examine faecal samples from 30 sheep for parasite egg counts, simple random sampling would require that each sheep is first identified and assigned a number between 1 and 600. If the animals are not already identified (e.g. with ear tags), this process may be time consuming or impractical. Using systematic sampling, prior identification of the animals is not necessary. The sheep are held in a yard, and allowed to pass, one by one, through a race. Every

20<sup>th</sup> animal is then selected as it passes through, giving a total sample of 30 animals.

Sampling interval

In this example, 20 is known as the *sampling interval*. It is calculated as  $N/n$ , or the population size divided by the desired sample size. Systematic sampling is only a form of probability sampling if the first animal selected is selected at random. In our example, a number between 1 and 20 would be selected at random, say 15. The 15<sup>th</sup> animal passing through the race would be the first animal selected, and then every 20<sup>th</sup> sheep. If however, the first animal is selected each time, there is no element of chance, and the technique becomes a form of non-probability sampling<sup>6</sup>.

The figure below illustrates different types of probability and non-probability sampling techniques to select 5 pigs from a herd. In the first, convenience sampling was used to pick the first pigs. In the second, purposive sampling chose the smaller pigs, as they are easier to handle. The third used random sampling, and the last uses systematic sampling with a sampling interval of 4 and a randomly selected first animal.

## Stratification

Stratified sampling

*Stratified sampling* is not a sampling technique, but an approach that may be used with any of the sampling techniques discussed. It involves dividing the population up into separate, exclusive groups (or strata), and selecting a sample from each group (stratum). If random sampling is used within each stratum, then it is known as stratified random sampling. Stratification can be based on any characteristic of the population. A survey of chicken farms may be stratified by production type, with broilers in one stratum and layers in another. A survey of a piggery may stratify by age, breaking the population into suckers, weaners, growers, sows and boars.

Reasons for stratification

There are three main reasons for using stratification. Firstly, it enables us to calculate estimates not only for the whole population, but for each stratum as well. For a national seroprevalence survey, stratification by province may provide a much more useful picture of the distribution of the disease. In this case, stratum estimates will be less precise than the combined estimate, because of the difference in sample sizes. The second reason for using stratification is that it is often operationally much more convenient. If stratifying by area, the survey can be done in stages, one area at a time, which distributes the work load more evenly. The third reason is because stratification may produce more precise results.

---

<sup>6</sup>Under most circumstances, random systematic sampling will produce a sample with very similar properties to simple random sampling. The same formulae as used for simple random sampling may be used to analyse the results from random systematic sampling. There is a danger, however, if the population has some sort of cyclic variation which matches or nearly matches the sampling interval. For example, let us consider a study to estimate the average temperature inside a poultry shed. Records are kept using an automatic recording device which records the temperature every hour. Records for 30 days are available (720 measurements). To save analysing all the data, a systematic sample of 30 measurements is taken (giving a sampling fraction of  $720/30=24$ ). A random number between 1 and 24 is selected, say 3, and the first record selected is the 3rd record. After that, every 24th record is selected. This sampling scheme would result in the selection of records which were taken at exactly the same time every day for the 30 days. If record keeping had commenced at midnight on the first day, then every record would represent the 2:00 am measurement for each day. Fortunately, such clear cyclic ordering in livestock surveys is very unusual, and so is rarely a problem.

**Example:** A survey is conducted to estimate the prevalence of insect-borne blood parasites in cattle, in a country with several different climatic zones - Zone A (hot and humid), Zone B (hot and dry), and Zone C (cool and wet). The insect vectors for the disease are likely to be much more common in the hot and humid part of the country (Zone A) than in the other two zones. A national survey will find a lot of variability in the prevalence from one area to another. Stratification by climatic zone means essentially that three separate surveys are conducted, one in each zone. In Zone A, a uniformly high prevalence is likely to be found, while a low prevalence is likely in the other two zones. In each survey, the variability is much less. When the results of the three surveys are combined, the overall variability of the results is less, so the precision is greater.

To achieve this precision, the aim is to have the animals within each stratum as similar as possible with respect to the characteristic in question, and for animals in the different strata to be as different from each other as possible (low within-stratum variability, and high between-stratum variability). Before a survey we rarely know enough about the distribution of the characteristic in the population to ensure this, but we may know that the characteristic is linked to some other factor (in our example, climatic zone). Stratifying on this factor will therefore help to increase the precision of the survey.

Examples of  
Non-Probability Sampling

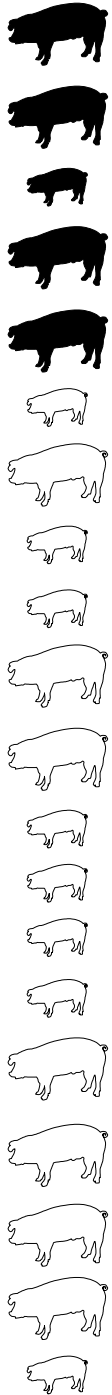
Examples of  
Probability Sampling

Convenience

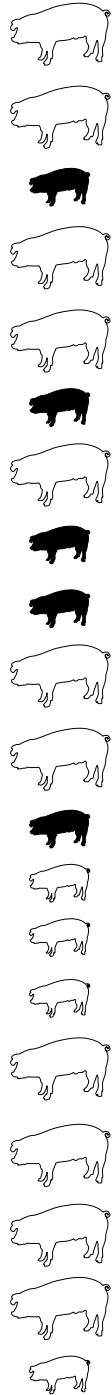
Purposive

Random

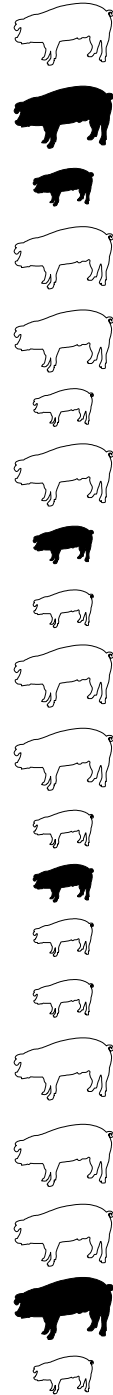
Systematic



First five pigs selected



Small pigs selected for ease



Random numbers used to select pigs



Second pig selected with random number, then every fourth pig selected

## Probability proportional to size sampling

In simple random sampling, every unit of interest in the population has the same chance or probability of being selected in the sample. Another probability sampling technique that is often useful is *probability proportional to size* (PPS) sampling. Instead of the chance for all units of interest being equal, in PPS sampling, the chance of a unit of interest being selected is proportional to some measure of the size of that unit of interest.

**Example:** A survey of villages in a large province uses PPS sampling to select the villages, based on their pig population. The population is all villages in the province and the unit of interest is the village. Records of the total number of pigs in each village are collected and used as a sampling frame. Villages with a large number of pigs have a higher chance of being selected than villages with a smaller pig population.

PPS sampling requires reliable information on the size of each unit of interest in the population. When this information is available, PPS can be used for very efficient survey designs. Unfortunately, it is often difficult to find this sort of information.



It is possible to do PPS sampling in a similar way to that described above for simple random sampling, but it is much more practical to use a computer to do the task. A program for this task, **Random Village**, is described on page 50.

Using the computer program requires that the sampling frame is available on computer disk. If this is not the case, then it may be necessary to do PPS sampling by hand. Use the following procedure:

Manual PPS sampling

**Step 1:** The sampling frame must be a list identifying all the units of interest in the population, with data about the size of each of them (usually livestock population of the herd or village). Add another column to this list for the *cumulative total*.

Cumulative total

**Step 2:** In the new column write down the cumulative total next to each item. The cumulative total is the size of the current item, plus the cumulative total from the previous line, as shown below.

Village Name (Unit of interest)	Cattle population (Size of unit of interest)	Cumulative total of cattle
Ban Dong	232	232
Ban Hai	89	321
Sisakhet	144	465
Si Meuang	129	594

**Step 3:** The last line in the cumulative total column represents the total number of animals in the entire study area. In PPS sampling, instead of picking a random number representing a village, we pick a random number representing an animal, and select the village that contains

the animal. Using any of the techniques described, pick a random number between 1 and the total number of *animals* in the population.

**Step 4:** Search down the cumulative total column until you find the last number which is equal to or greater than the randomly selected number. The unit of interest containing that number is the element selected.

**Example:** If the above list represents the entire population of villages in the study area, then the total number of animals in the four villages is 594. A number between 1 and 594 is selected at random, say 256. Searching down the list, the second village contains the number 256, and this village would be included in the sample.

**Step 5:** Continue until the required number of units of interest has been selected.

## Sampling frames

Sampling frame

In random sampling, every unit of interest in the population has the same chance of being selected. In the techniques described above, this is achieved by using random numbers, and picking units of interest from a list. This list is called the *sampling frame*, and should contain every unit of interest in the population.

**Example:** A survey is conducted in a large intensive piggery to estimate the prevalence of pigs with respiratory disease. The farmer keeps a list of all mature sows and boars, each identified with an ear tag number, so this list is used as the sampling frame. Twenty animals are selected using random numbers from the sampling frame, and these animals are examined for signs of respiratory disease.

Clearly this survey has a problem with selection bias. It is not possible to infer the true prevalence in the piggery from the survey results, because the selection bias means that the sample is not representative of the population. The sampling frame does not include grower pigs, only mature breeding pigs. As respiratory problems are more common in growers than adults, we are likely to get misleading results, even though we used random sampling. This is because the sampling frame was incomplete, and did not include every animal.

A different problem can occur when a sampling frame lists the same units of interest more than once. In a village sampling frame, if one village is listed twice, then it has twice the chance of being picked as other villages. Another difficulty that can happen is the problem of identifying the elements from the list. Sometimes there may be two animals with the same ear-tag number, or two villages with the same name. The ideal sampling frame is therefore a list which:

A good sampling frame

- contains *every* unit of interest in the population (no omissions);
- contains every unit of interest *only once* (no duplications); and
- *uniquely* identifies every unit of interest.

Sampling frames may also contain other information to help with more complex sampling schemes. One example is a village sampling frame that lists all villages in an area, but also includes information on the livestock population of those villages

(the size of the village). This extra information can be used for probability proportional to size (PPS) sampling described earlier on page 48.

Sources of possible sampling frames

Sometimes a suitable sampling frame already exists. When surveying villages, the government statistics office, and many other government departments, usually maintain lists of villages, often computerised with unique identification numbers. These lists are very useful as a sampling frame. The statistics office or agriculture department may also maintain information on livestock population, but to be useful, this needs to be up-to-date, and have the population of each village instead of summary figures for districts or provinces.

Farm sampling frames

Sampling frames for farms may be harder to find. Registration records for commercial farms, or farm supply companies (e.g. feed suppliers), may be able to provide information, but it is important to check how complete this is.

Any sampling frame is likely to be imperfect, either missing a few members of the population, or not identifying others properly. Just because a sampling frame isn't perfect, it doesn't mean it can't be used. It is a matter of judgment to assess how good the sampling frame is, and whether the problems with it are likely to affect the results of the survey. For instance, if a sampling frame is missing 20% of the population, it may be better to try to find a better sampling frame. However, if there is no pattern as to which members are missing, then results from a survey using the frame could be perfectly adequate. On the other hand, even if there is only a small number missing from the frame (say 5% or 10%), but there is a clear pattern (for instance all the biggest farms are missing), then there is a significant danger of the results being biased.

If there is no sampling frame

In many cases, no existing sampling frame is available. To carry out a survey using random sampling, it is then necessary to either:

- build a new sampling frame, by identifying all the units of interest in the population and creating a list;
- use a different sampling strategy that requires a different type of sampling frame which is easier to obtain (such as two-stage sampling, described on page 64); or
- use a specialised technique for random sampling with no sampling frame (random geographic coordinate sampling, described on page 65).

## Selecting a sample from a sampling frame

When a good sampling frame is available or has been constructed especially for the survey, the method of selecting elements from the sampling frame is quite simple, and has been described above under Random Sampling Techniques (page 40). A random number table or a computer can be used to pick random numbers, and these are used to identify members of the population.

When the sampling frame is available on computer, this job can be made much faster and simpler, by using a specialised program.

The Survey Toolbox includes one program, **Random Village**, to do this job. The program, shown below, was designed to pick villages at random from a computerised list, but can be used to pick anything (farms, regions, districts, animals, feed supply shops) as long as there is a computerised sampling frame listing all the members of the population. Chapter 6 has more information about computerised lists (databases) and computer management of information.



### Requirements

To use a database file containing the sampling frame for random selection, the file must be in dBASE, Paradox, or ASCII format. The file can contain any number of fields, but must contain at least one field with a unique identifier (name or ID number) for each element. Optionally, it can also contain:

- a field to be used for stratification (for example, district for a village sampling frame, or enterprise type for a farm sampling frame), again as a name or number;
- a size field to be used for probability proportional to size sampling (PPS). This field must be a whole number (integer).

### Epi Info

If building your own sampling frame, you can use Epi Info to enter and manage the data. Creating a new table using Epi Info is described on page 130. When creating the table you need to create a data entry screen with field codes. An example screen suitable for creating a village sampling frame is shown on the next page.

Demonstration Data Entry Form	
Village Sampling Frame	
Village ID:	#####
Village Name:	_____
District ID:	##### (Stratification field, if used)
Population:	##### (Size field for PPS sampling)

Using an existing database or table as a sampling frame

### Selecting villages

To start the **Random Village** program, use the Windows **Start** Programs menu to bring up the list of programs, select **Survey Toolbox**, and choose **Random Village**. Alternatively, you can use either the **Survey Toolbox** main menu.

When the program is running, select random villages by following these steps:

- Step 1:** Click on the **Open** button to open the database containing the sampling frame. This will open the Open File dialog where you can change directories or drives to find the file you want. Select the file, and click on the **Open** button.
- Step 2:** First you need to specify how many elements you want to select (the sample size). Enter the sample size in the Number to Select box, or use the arrows to change the number up or down. Part II explains how to work out the sample size you need.
- Step 3:** Select the sampling type. If using simple random sampling, you can leave it as it is. If you are using PPS (probability proportional to size sampling, see page 48), select Probability Proportional to Size from the Sampling Type box. You then need to specify which field holds the information about the size of each element. Click on the arrow at the right of the Size Field box, and select the field with the size information (e.g. village livestock population).
- Step 4:** Now you need to specify whether to use replacement or not (See page 44). Choose With Replacement, or Without Replacement in the Replacement box, depending on the survey design. Usually you don't have to change this (leave it as With Replacement).
- Step 5:** Next, specify if you want to use stratification. Stratification was described on page 45. If you want the sample stratified, click on the check box, and select the field that has the information for stratification.
- Step 6:** Then, you need to tell the computer what information you want displayed. A list of all the fields in the database is shown on the right under Identification Fields. Choose the field that contains the unique identifier. By holding down the Shift key, and clicking with the mouse, you can choose a range of fields, and using the Control key and the mouse selects several fields. All the fields that are selected will be listed for the random elements, so pick all the fields that help you identify the element.

- Step 7:** Finally, you can choose to select from only a part of the list in the file. For instance, if the file contains a list of all villages in the country, but you wish to conduct a survey only in one province, you can instruct the program only to work with villages in that province. First click on the Sample from Subgroup checkbox. Then, under Group Field, pick the field that contains the grouping information. In this example, you would chose the field with the province identifier. Next chose the relationship. You can define a group as all elements that are equal to, greater than or less than the value specified. In this example, we would chose "is equal to" from the list. Lastly, enter the value under Group Identifier. In this example, you could pick the province name or number from the list to specify which province you want. As another example, if you wanted to survey only villages with cattle, you could specify Cattle as the Group Field, "is greater than" as the relationship, and type in "0" as the Group Identifier. Only villages with cattle would be included in the sample.
- Step 8:** When all the settings have been made, click on the **Select** button, to have the computer chose the random elements. A window will open listing all the selected elements. You can click on the **Print** button to print the information, or else the **Save** button to write the information to an ASCII file or a new database.

The screenshot shows a window titled "Random Village" with a sub-header "Survey Toolbox". It contains a table with the following columns: No, VILLAGEID, VILLAGENAME, DISTRICTNO, CATTLE, and BUFFALO. The table lists 18 villages with their respective IDs, names, district numbers, and counts of cattle and buffalo. Below the table are five buttons: Font, Print, Save, Select Another, and Exit.

No	VILLAGEID	VILLAGENAME	DISTRICTNO	CATTLE	BUFFALO
1	145	Xankao	3	67	88
2	158	Xoknoy	3	154	236
3	150	Nonnakbay	3	149	124
4	159	Xokyuai	3	215	226
5	155	Nakhuytai	3	115	101
6	107	Nongtaenuane	2	195	83
7	109	Cheapa	2	302	39
8	111	Nongboaktai	2	132	118
9	115	Xankat	2	77	39
10	98	Tadthong	2	214	63
11	52	Thinethen	7	78	40
12	54	Sinaanomuane	7	621	362
13	45	Thintom	7	215	311
14	46	Xaikhuane	7	125	423
15	35	Nonghai	7	120	100
16	336	Senot	8	144	156
17	349	Nanieng	8	109	149
18	338	Khokpheung	8	97	190

Enter data for a new sampling frame

There are several other options within the program. Instead of opening an existing database, you can click the **New** button to create a new database file for your sampling frame. You then need to enter the information into the computer yourself, before you can select random elements.

Replacing selected elements

If you do have a database with your sampling frame, but there are some mistakes in it, the **Edit** button opens the file ready for editing. You can change any information, and then return to the main screen to select random elements.

When random elements have been selected, there are two other buttons under the list: **Font** and **Select Another**. The **Font** button allows you to choose a different font to display the list. This is useful if the database has the names of villages listed in a non-English script.

To use the **Select Another** button, click on one of the randomly selected elements in the list, and then click the Select Another button. This will delete the selected element and randomly select a new one.

**Warning:** this should only be used when absolutely necessary. The only reason for using this button is when it is impossible to survey that element. In a village survey, this may be because the village is extremely inaccessible or travel to it is dangerous. The problem with using this button is that the sample is no longer random, because each village doesn't have the same chance of being selected. Using this button to select new elements just because one is inconvenient can lead to invalid results, meaning that all the effort that was put into the survey was wasted.

## Selecting animals from a village population

A common problem encountered in livestock disease surveys in developing countries is deciding how to select a random sample of animals from within a village. In many countries, the smallholder village farming system is an important part of the livestock industries. One village usually has many livestock owners, each with varying numbers of animals. Often, these animals are in relatively close contact, either roaming unrestrained (e.g. village chickens) or housed or grazing together. Because of the close contact between animals belonging to different owners, it is easy for contagious diseases to spread through the village. Usually, from the point of view of a disease survey, all the animals in the village can be considered to belong to one large herd, even if they are owned by many different owners. All animals are generally exposed to the same diseases, and are reared using similar husbandry techniques.

When conducting livestock disease surveys of smallholder livestock raised in the village system, it is usually more sensible to treat them as a single herd, and to draw a simple random sample from the population of all animals in the village. This is difficult, as the animals are owned by many different people, and there is usually no sampling frame available. Even if figures on the village livestock population are recorded, these are usually only collected once per year, and are often too out of date to be of any use. In addition, animals are rarely individually identified (e.g. with ear tag numbers).

To overcome these problems, a practical technique for randomly selecting individual animals in the village is described below. To illustrate the technique we will use an example:

**Example:** A survey is conducted in a village to monitor the effectiveness of a haemorrhagic septicaemia vaccination program. The aim is to estimate the prevalence of cattle and buffalo in the village with antibodies against haemorrhagic septicaemia. As both species are susceptible, and animals in the village are in close contact, all the cattle and buffalo are treated as a single herd. The village has 48 households that raise either cattle or buffalo or both. There are 49 buffaloes and 125 cattle in the village for a total of 174 animals, and a random sample of 20 animals is required.

## Building the sampling frame

Village interview

The first task is to build a sampling frame, listing and uniquely identifying every animal in the village. It is unlikely that any one person in the village would know exactly how many animals each of the 48 households had, and any records are likely to be out of date. One option is to walk around the village and either ask the owners, or count the animals directly (that is, conduct a census of animals in the village). Conducting a census is time consuming and it is easy to miss some animals, but it may be the best approach in some circumstances. Another approach that is sometimes useful is to hold a *village interview*, to which all the livestock owners are invited, and to ask them how many animals they have. A village interview with all the livestock owners requires some organisation, and may take a few hours to complete, but if this is possible, it is an efficient way to collect information for a sampling frame.

Village interviews are also extremely useful for collecting different types of information. These are discussed in detail in Chapter 5, along with guidelines on how to run an interview. The collection of information for a sampling frame is also discussed there (see page 115), so only a brief description will be given at this stage.

It is important to try to get as many of the village livestock owners as possible to attend the meeting, to make it easier to build a complete sampling frame. After explaining the purpose of the survey, each livestock owner present at the meeting is asked, in turn, what their name is and how many animals they keep. This information is recorded on a sheet with the columns shown below (A full copy of an example data recording sheet is reproduced in Appendix D.)

N°	Name	Cattle	Buffalo	Total	Cum. Total	Selected
1	Lung Noi	5	-	5		
2	Tu Nyai	2	3	5		
3	Silipak	-	4	4		
4	Khamphone	8	2	10		

When the information has been collected from each livestock owner present at the meeting, collect information about those livestock owners *not* present. The group is asked to identify all the owners who are not at the meeting, and to estimate how many animals they keep. This step may take some persistent questioning, and require prompts to help the owners think of others not at the meeting. Experience

has shown that village interviews are usually able to make a list that contains almost every animal in the village.

### Selecting the “number” of the animals

The list completed during the village interview serves as the sampling frame, but is different to the sampling frames discussed earlier. When drawing a random sample of animals, a sampling frame will usually be a list of all animals, with an identification number. In this case, the list is of all livestock owners (identified by name and a line number), with the number of animals kept. This list may be used as an animal sampling frame (rather than an owner sampling frame) because each animal in the village appears on it (although they are not yet individually identified – we will solve this problem later).

Random selection of animals

The list can now be used to randomly select animals. There are two ways of doing this: using a random number table, or using a computer. The computer technique is slightly faster and simpler, but requires a notebook computer to be available in the village during the survey. As this is not often possible, the manual technique using a random number table is described first.



#### Random number table

The technique for selecting random animals is slightly different to that described previously (page 41) because the sampling frame is different. It is similar to the technique used for probability proportional to size sampling (page 48). To pick animals use the following procedure:

**Step 1:** On the data recording sheet, calculate the cumulative total number of animals and write it in the column marked Cum. Total. The cumulative total is the total number of animals kept by all livestock owners in the village up to that point.

**Example:** The cumulative total for owner N° 1 is just the total number of animals, 5. The cumulative total after owner N° 2 is equal to the number of animals kept by owner N° 2 (5), plus the previous cumulative total (5), which equals 10. The cumulative total after owner N° 3 is 4 plus the previous cumulative total (10), or 14. This is continued to the last owner. Note that the last number is equal to the total number of animals in the village.

N°	Owner Name	Cattle	Buffalo	Total	Cum. Total	Random Number	Animal Selected
1	Lung Noi	5	-	5	5		
2	Tu Nyai	2	3	5	10		
3	Silipak	-	4	4	14		
4	Khamphone	8	2	10	24		

The numbers in the cumulative total column represent ID numbers for each animal in the village. Owner N° 1 has animals with ID number 1 to 5. Owner N° 2 has animals with ID numbers 6 to 10, and so on. These new animal ID numbers can now be used for random sampling.

N° 1 (Lung Noi)					N° 2 (Tu Nyai)					N° 3 (Silipak)				N° 4 (Khampone)									
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

**Step 2:** Using a random number table pick the first random number (see page 42 for instructions on using a random number table to select random numbers). The random number must be between 1 and the total number of animals in the village, given by the last number in the cumulative total column. This number represents an animal that is to be selected. Find the owner of the animal on the list.

**Example:** In our example village, we need to pick a number between 1 and 174 (the total number of cattle and buffalo). If we picked the number 12, we need to identify who owns animal number 12. Search down the cumulative total column for the first number greater than 12, which is 14, at owner N° 3. This means that owner N° 3 is the owner of animal number 12.

**Step 3:** Now we have identified the owner, we need a way of identifying the individual animal. We need to calculate which animal in numeric order the selected animal is.

**Example:** We have selected animal number 12 belonging to owner N° 3. Owner N° 3 has 4 animals, and we need to decide which animal we want. The animals belonging to owner N° 3 would be numbered 11, 12, 13, and 14 if we had a true list of individual animals. If we want animal number 12, it is the *second* animal belonging to owner N° 3. A quick way to calculate this is to subtract the cumulative total for the previous owner from the random number selected. In this case, we would take 10 (the cumulative total for owner N° 2) from 12 (our random number) to give 2. This means that we want the second animal belonging to N° 3.

**Step 4:** Record the number of the animal next to the owner in the Selected column. Then repeat steps 2 and 3, selecting more random numbers from the random number table and finding the animal in the same way. Continue until enough animals have been selected. If the same animal is selected twice, discard that random number and pick a new one. This is because we always use sampling without replacement when selecting animals. It is possible to pick several animals belonging to the same owner.

**Example:** Three more random numbers are selected: 17, 3 and 20. The animals have been selected and recorded on the sheet below. Check for yourself to see how it was done.

N°	Owner Name	Cattle	Buffalo	Total	Cum. Total	Random Number	Animal Selected
1	Lung Noi	5	-	5	5	3	3
2	Tu Nyai	2	3	5	10		
3	Silipak	-	4	4	14	12	2
4	Khamphone	8	2	10	24	17, 20	3, 6

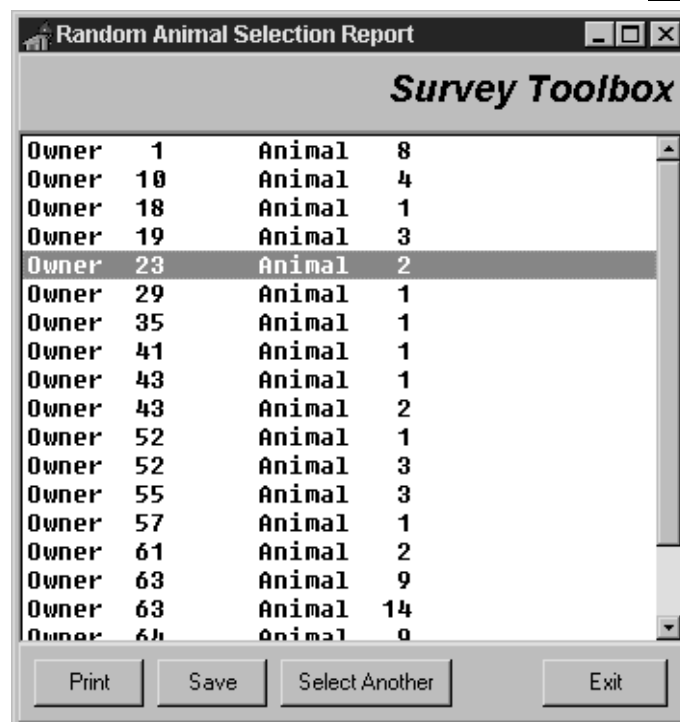
### Computer program



The same procedure can be automated using one of the Survey Toolbox programs included with this book. The program is called **Random Animal** and can be started using the Windows **Start** menu, selecting Programs, then Survey Toolbox, and Random Animal.

- Step 1:** Collect the information on the number of livestock kept by each livestock owner in the village in the same way, and record it on the data recording sheet. (There is no need to calculate cumulative totals).
- Step 2:** Click in the Animals column of the box on the left of the screen, and type in the number of animals owned by Owner number 1. Press the **Enter** key or down arrow to move to the next line. The owner number is automatically entered for you.
- Step 3:** Continue entering the number of animals kept by each livestock owner, pressing the down arrow to move to the next. Make sure that the owner number on the screen corresponds to the owner number on the form, as this will be used to identify the owner.

- Step 4:** If you make a mistake, you can go back and change it. You can use the buttons at the bottom of the screen to add or delete records, or move up or down the list.
- Step 5:** When all owners have been entered, click on the Select buttons to indicate whether you want to select a fixed number of animals or a fixed proportion of the village population. This depends on the survey design being used (see Chapter 7).
- Step 6:** Enter the number of animals or percent of the village population you want to select in the Number to Select box. You can type the number in or use the up and down arrows to change it.
- Step 7:** Click the **Select** button to randomly select the animals.
- Step 8:** The report window opens listing all the animals, and their owner numbers that you need to select for the sample. You can print the list to a printer, or save it to disk with the **Print** or **Save** buttons.



- Step 9:** Use the **Select Another** button to choose a replacement for an animal selected from the list.

**Warning:** Use this button only when absolutely necessary. You may need to select another animal when the selected animal cannot be captured. However, you should make every effort to sample all the animals selected and use this button as rarely as possible. When you select a replacement animal, the sample is no longer random, and the results may be biased.

## Identifying selected animals

Regardless of which method is used to randomly select animals, the result will be a list of owners and animal numbers that looks like this:

Owner 1 Animal 3  
 Owner 3 Animal 2  
 Owner 4 Animals 3, 6  
 Owner 8 Animal 7  
 Owner 11 Animals 2, 9, 16  
 Owner 22 Animal 2

The approach to identifying which individual animals should be included in the sample based on these numbers is as follows:

- Step 1:** Identify the owners of the selected animals. If selection was done manually using a random number table, the owners' names are already written on the list. If selection was done using a computer, you will need to copy the information back onto the data collection sheet, using the owner number to identify the correct owner.
- Step 2:** Random selection can take place during the village interview. At the end of the village interview, all selected owners should be asked for their permission to allow the survey team to examine or collect specimens from their animals. Selected owners should be told that *all* the animals will have to be seen, even if only one or two are required for examination.
- Step 3:** For each owner, visit the place where their animals are kept. Look at the animals first without disturbing them.
- Step 4:** Ask the owner to count, out loud, all of the animals. In this way, the owner is assigning a temporary identification number to each animal. Check on the list for which animal or animals are required, and note which ones are assigned the relevant numbers.



**Example:** Using the list shown above, for owner 4, animals number 3 and 6 are required. On visiting the place where the animals are kept, there are 8 cattle and 2 buffaloes. The owner starts counting the animals out loud, starting with the buffaloes as 1 and 2, then the cattle. The first of the cattle counted is number 3, which is one of the animals required. As counting continues, another of the cattle is given the number 6, which is also required. When the owner has finished, they are asked to catch the two cattle given numbers 3 and 6 for examination or specimen collection.

**Step 5:** The process is repeated for each of the selected owners until all animals have been examined.

The technique can be used in a wide variety of situations, though some modification may sometimes be necessary. It may seem complex initially, but with good planning, and good communication with farmers it is simple to carry out. A few points need special attention.

- It should be explained to the owner that the survey team needs to look at the animals before capturing any. This is to avoid the problem of the owner selecting and capturing animals before the survey team arrives, thereby invalidating the random sample.
- Anybody can count the animals to assign numbers, including a member of the survey team, but it is very important that the person counting *doesn't know which numbers are to be included in the sample*. If they know, they could (consciously or subconsciously) select animals which are easier to restrain.
- Sometimes livestock owners keep their animals in separate groups, rather than all in the one place. Before visiting the animals, ask the owner how many animals are in each group. The groups can then be assigned ranges of numbers to help decide which needs to be visited.

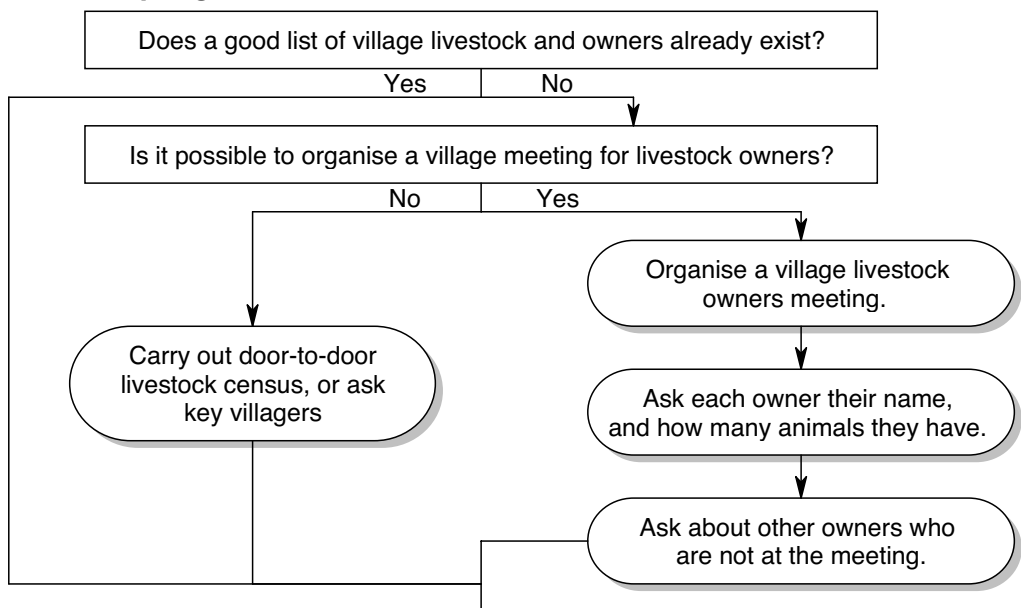
**Example:** Survey staff talk with a cattle owner before visiting her animals. She says that she keeps 24 animals in three groups. The first group of 7 animals is kept a short way from the village. This group is assigned numbers 1 to 7. The second group of 12 animals is kept at the owner's house in the village. These are assigned numbers 8 to 19. The other animals are at her brother's house. These are given numbers 20 to 24. If animal number 14 is required, then only the group containing that animal (the ones at the house) need be visited. These are counted, starting at number 8. When using this system, it is important that the person assigning numbers to the groups does not know which animals are to be selected.

- Sometimes animals cannot be examined or specimens collected. This may be because they cannot be captured, restraint equipment fails and they escape, they are too far from the village to reach, or the animal is agitated and dangerous to restrain. In these situations, a different animal will have to be substituted. When an animal is replaced, the sample is no longer completely random, because the chance of the replaced animal being in the sample is zero. This should therefore be avoided whenever possible. However, for practical reasons some animals will need to be replaced. When selecting a new animal, the same

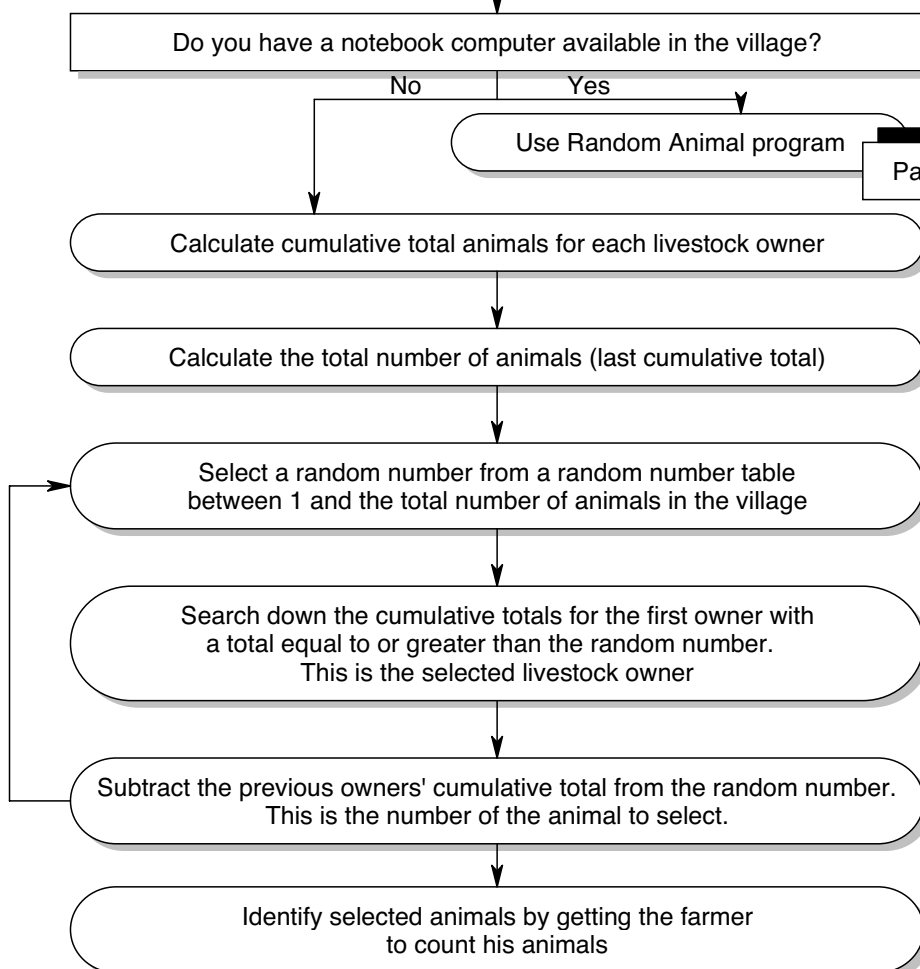
procedure as that described above should be used, using either a random number table or a computer. It is better not to just pick another animal belonging to the same owner.

### Sampling Animals in the Village

#### 1) Build sampling frame



#### 2) Select animals



Page 58

## Two-stage sampling

When selecting a sample using simple random sampling, all animals need to be first identified and listed on a sampling frame. In systematic sampling, they need to be "lined up" in some sort of sequence for sampling. When surveying very large populations (e.g. national level surveys) both simple random and systematic sampling are impractical, if not impossible. For example, it is not possible to create a list containing every single chicken in a country with a total population of 40 million chickens.

*Two-stage sampling* addresses this problem by gathering the animals into convenient groups. While a list of all chickens would be impossible to compile, a list of all chicken *farms* in the country may be more easily obtained (from registration records, for instance). In two-stage sampling, groups of animals are selected first (e.g. a number of chicken farms) and then individual animals are selected from the selected groups. At the first stage, the population is all the chicken farms in the country and the unit of interest is the farm. At the second stage, the population is all the chickens on each selected farm, and the unit of interest is the chicken. At each stage of two-stage sampling, the sample is selected by random sampling. Stratification may also be used, usually at the first stage (for instance, stratifying farms by production type, size or district).

Advantages of two-stage  
sampling

Surveys using two-stage sampling have two distinct advantages: they are easier to plan, as a complete list of all animals in the study area is not needed, only a list of the first-stage units; and they are more practical for the field work team, as fewer sites need to be visited. The disadvantages are that the results may not be as precise as with simple random sampling, and the formulae for analysis of the data can be very complex.

**Example:** A survey is planned to assess the impact of hog cholera on smallholder pigs. The population of interest is all the village pigs in the country, a total of 5.5 million. It is not possible to create an accurate sampling frame listing all these animals. However, a list of all the 18,322 villages in the country does exist, maintained by the National Statistics Office. This list is available on computer disk. A two-stage survey is therefore used. At the first stage, 40 villages (the first-stage units of interest) are selected from the village sampling frame, using a computer. At the second stage, 10 pigs (the second-stage units of interest) are selected at random from each of these 40 villages, to give a total sample size of 400 pigs. As no list of pigs is available for the villages, a village interview is used to create a sampling frame in each village.

Two-stage sampling is commonly used in livestock surveys. Chapter 7 provides details of the design, implementation and analysis of data for a two-stage prevalence survey.

## Sampling without a sampling frame

Sometimes it is not possible to use a sampling frame for sampling, because no sampling frame exists and it is not practical to create one.

**Example:** A survey of villages is planned in a remote region that has recently suffered from political and social instability. Government infrastructure is weak, and large movements of people have taken place over the last few years. As a result, whole villages have disappeared, and new villages have appeared, but no records or maps exist of where these villages are, or how many there are.

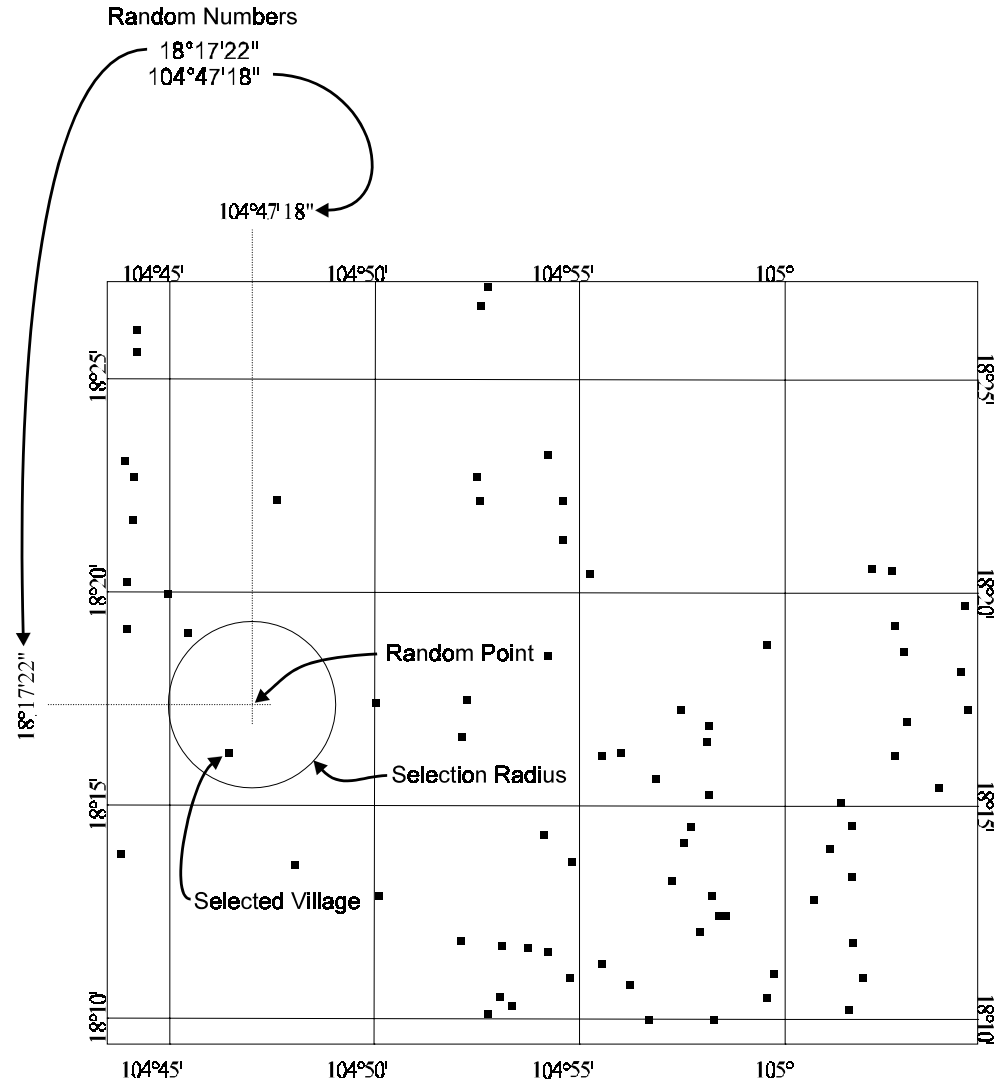
**Example:** A survey to measure the prevalence of animals with antibodies to brucellosis in nomadic goat herds is planned. There is no record of these herds, and they move to a new location every few days.

Random geographic coordinate  
sampling

In circumstances such as these, creating a sampling frame is virtually impossible, so another approach must be used for random sampling. *Random geographic coordinate sampling* (RGCS) offers a technique for the selection of a random sample without the need for a sampling frame. In RGCS two random numbers are selected, which are the x and y coordinate of a random point, somewhere in the study area. Towns, villages or herds that are located within a certain distance (the *selection radius*) from this random point are identified. If there is more than one town, village or herd near the point, one is selected at random. When the survey is completed and estimates are calculated, data from the selected herds are *weighted* proportional to the total number of herds within the selection radius of that point.

**Weighting** means that some  
data contributes more  
information than other data

**Example:** A sample of villages is being selected using RGCS for a survey to estimate the prevalence of Newcastle disease in village chickens. Two random numbers are chosen using a random number table, representing the coordinates of a random point. A selection radius of 2 km has been chosen. The survey team travel to that random point, then search for villages within 2 km of the point. They find that there are 3 villages within the selection radius. In order to decide which of the 3 villages to include in the sample, they select one at random from the 3. The selected village is then visited and the chickens examined. When the survey is finished, the results from the village are weighted by a factor of 3, the number of villages in the selection radius. This is because, in the survey, the selected village is representing all the villages in that area.



RGCS is useful mainly for the random selection of groups of animals, such as villages or herds, as these are large, relatively easily to identify, and can't change location very quickly (even migrating herds can usually be identified with a particular location on a particular day). It is not suitable for selecting individual animals which may move from place to place, but could be used, for example, to select families from within a village, based on the position of their house.

The problem with RGCS is that it involves much more work than selecting a sample from a sampling frame. This is because it is necessary to go out into the field to look for the herds or villages, rather than being able to select them before the field work. Some of the main difficulties include:

Problems with RGCS

- many randomly selected points will contain no villages within the specified distance, so new points need to be selected;
- locating and counting the villages requires much more travel than traditional surveys, so field costs are significantly higher; and
- identification of the location of the random point, and knowing whether a nearby village is within the specified distance may be difficult in the field.

These problems and the extra work involved mean that RGCS should be used only as a last resort. If there is any practical way of building a sampling frame, then

it is likely to be simpler than RGCS. For instance, if maps of the area exist, and these maps contain reliable information on the location of all villages, then the map can be used to list the names of villages and create a sampling frame. RGCS need not be used. If, however, the map does not show village locations, or only shows some of them, then any sampling frame built from the map will be incomplete. In this situation, RGCS offers the only means of selecting a random sample.

Global positioning system  
(GPS)

The process of selecting a sample using RGCS is described in detail below. There are two ways to make the task easier and more practical. The first is to use a Global Positioning System unit (GPS). A GPS unit is a hand-held device that uses a network of satellites to pinpoint its geographic location. The coordinates are displayed on a screen. A GPS unit is virtually essential for RGCS, and units capable of providing the level of accuracy normally required are relatively inexpensive (US\$500 - \$1000).



Geographical Information  
System (GIS)

The second technique is the use of specialised computer software for handling maps (a Geographical Information System, or GIS) and remotely sensed data (satellite images or aerial photographs). RGCS can be successfully carried out without a GIS and remotely sensed data, but if available, the task will be simpler, quicker and more efficient. The use of a GIS to assist with RGCS is described on pages 69 and 72.

Four-wheel drive vehicle

A further requirement of the technique in most cases is the use of a four-wheel drive vehicle for the field work. Random points may be in very inaccessible areas, and a four-wheel drive vehicle is usually needed to reach them.

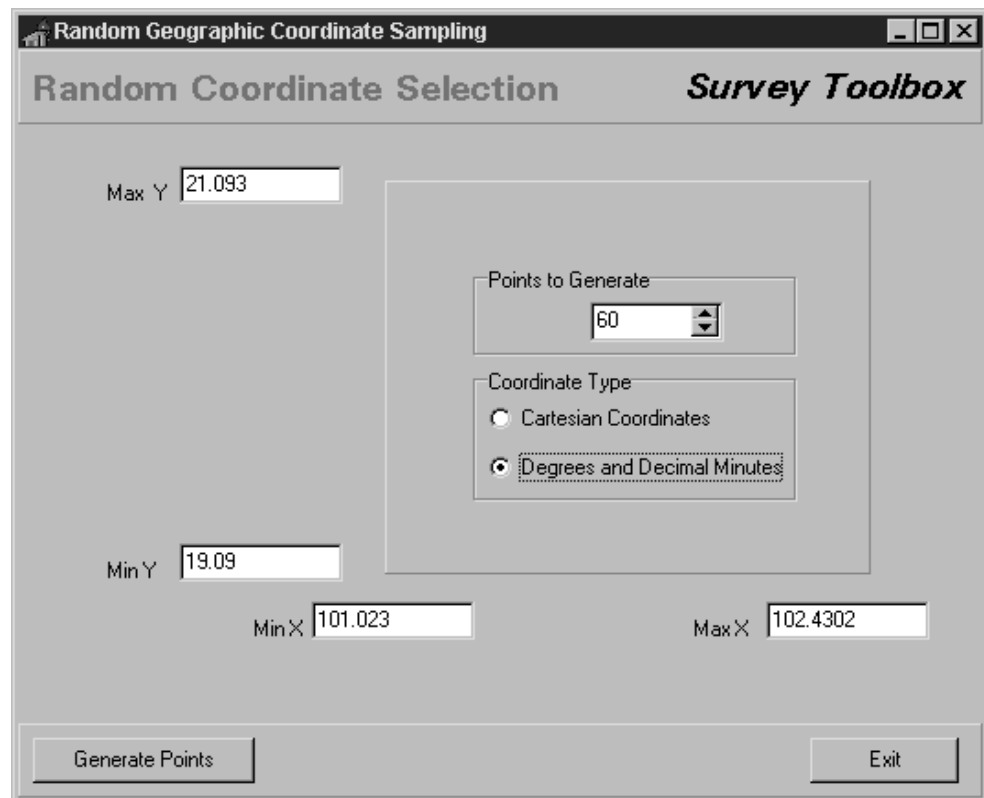
Traditional random sampling using a sampling frame is done in two parts, first selecting random numbers, and then identifying the elements to be selected based on those numbers (using the sampling frame). Similarly, RGCS involves first the selection of random coordinates, and then a process of identifying members of the population to be selected based on those coordinates.

## Selecting random coordinates

Random coordinates are simply pairs of random numbers that represent points within the study area. They can be selected using a random number table (as described on page 42) but are more conveniently selected using a computer. Two programs are included in the Survey Toolbox for the selection of random points, one of which works with the ArcView GIS program.

### RGCS for Windows 95

The **RGCS** program that is included in the Survey Toolbox creates a number of random coordinates within a rectangular area. To start the program use the Windows Start menu, select Programs, then Survey Toolbox, then RGCS.



To select random points:

**Step 1:** Enter the boundary coordinates of the study area. You will need a map with a coordinate grid to find these figures. Max Y is the upper y coordinate, or the northernmost latitude of the study area. Min Y is the lower y coordinate or the southernmost latitude. Min X is the left x coordinate, or the westernmost longitude. Max X is the right x coordinate, or the easternmost longitude. If the figures on the coordinate grid on your map are in degrees and minutes, convert them to decimal degrees before entering, by dividing the minutes by 60 and adding to the degrees.

**Example:** A survey is planned for a single state. A map of the state is obtained, with a coordinate grid marked in degrees and minutes. The

latitude of the northernmost point in the state is read from the map as 14° 34'. To convert this to decimal degrees  $14^{\circ} 34' = 14 + (34/60) = 14.5667$ . This number is entered into the Max Y box. The procedure is repeated for the other boundaries.

- Step 2:** Specify how many random points to select. When selecting from a sampling frame, each random number corresponds to a single member of the population, so the number of random numbers needed is equal to the sample size. With RGCS, some random points will have no herds within the selection radius, and have to be discarded. Some of the random points will also fall outside the study area. For these reasons, you should select more random points than your sample size. The number of points should be equal to 2 or 3 times the sample size. Enter the number required in the Points to Generate box, or use the arrows to change the number.
- Step 3:** Specify how you want the results to be displayed in the Coordinate Type box. Cartesian coordinates are normal x, y coordinates, that may represent metres, kilometres, feet, yards or miles on a map grid. Degrees and decimal minutes are the latitude and longitude grids. Click on the button that corresponds to the units your map is marked in.
- Step 4:** Click on the **Generate Points** button. A numbered list of points is displayed.
- Step 5:** Discard those points that fall outside the study area. Check each point in order against the map. Identify the location of each point, and check if it is inside the study area or not. If not, discard it using the **Delete** button.
- Step 6:** Print or save the list using the **Print** or **Save** buttons.

### RGCS for ArcView GIS v.3

Also included in the Survey Toolbox is a random point selection program, **RGCS ArcView**, designed to work with ArcView GIS version 3<sup>7</sup>. Those users who have a copy of this program are able to load the extension file, and have ArcView select points within a specified area automatically. In order to use this program, you must have a copy of ArcView, and you need a digital map (theme or coverage) of the study area, in a format that ArcView can read. If you don't have access to ArcView, skip to the next section, Field Procedures on page 74.

To install the extension, copy all the files from the CD directory \Survey Toolbox\AVRGCS to the extensions subdirectory for your copy of ArcView (usually c:\ESRI\Av\_gis30\ArcView\Ext32).

Follow these steps to load the extension and select random points:

- Step 1:** Start ArcView. With the Project window active, select the File menu and choose Extensions.
- Step 2:** In the Extensions dialog search down the list until you find "Random Geographic Coordinate Sampling". Click on the checkbox on the left

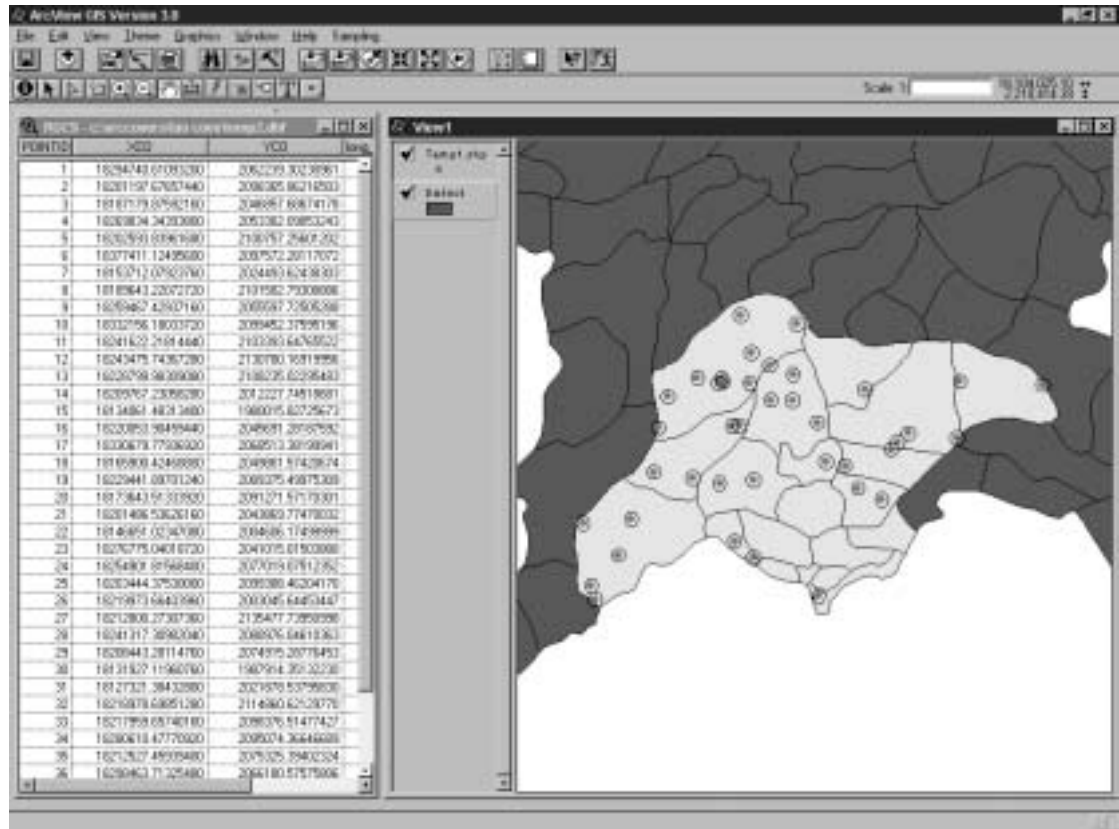



---

<sup>7</sup>ArcView GIS version 3, ©1992-1997, Environmental Systems Research Institute, Inc., 380 New York Street, Redlands, CA 92373 USA.

to select it, then click OK to load the extension. An introductory screen and brief instructions will appear.

- Step 3:** Create a new View. Add a new Theme, and load the digital map of the study area into the theme. The map may be of an area larger than the study area, but must have one or more polygons which describe the study area. For example, if conducting a survey of one province, a national map without provincial boundaries is not adequate, but a national map showing the boundaries of all provinces can be used. A map showing all districts within the province can also be used.
- Step 4:** Click on the Select tool, and select the polygons which represent the study area. This may be a single polygon (e.g. one province on a national map), or several separate polygons (e.g. all the districts making up one province).
- Step 5:** To start selecting points, open the Sampling menu and chose Select Random Points. Alternatively, you can click on the **Run** button, on the right of the button bar.
- Step 6:** The program asks how many points to select. Enter a number 2 or 3 times the sample size, to account for points with no nearby herds.
- Step 7:** The program asks you to specify the selection radius. Enter the desired selection radius in *map units*. This means if distances on the digital map are measured in metres, and you want a 2km selection radius, enter 2000. If the digital map is in geographic coordinates (degrees and minutes) you will need to convert the distance to degrees. One kilometre is approximately equal to 0.009 degrees (north-south).
- Step 8:** The program then asks for a file name to store the random points, first as a point theme, and then as a file in dBASE format. These files are stored in the default ArcView directory.
- Step 9:** Finally, the program selects the points within the selected study area, and displays them on the map. Each point is surrounded by a circle defined by the selection radius. The database file of coordinates is also displayed and may be printed.



Once the random points have been selected, nearby villages need to be identified, and sample villages selected. This process is described in the next section. GIS software and remotely sensed data can be used to make the field work simpler and more efficient, as described on page 72.

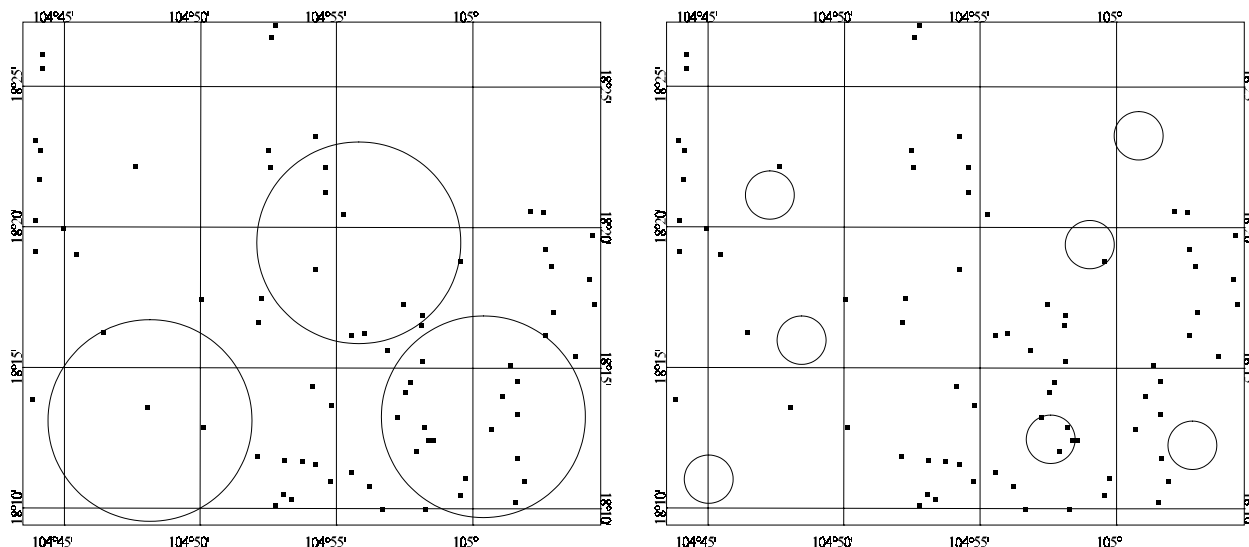
## Identifying selected villages

The only reliable way to identify villages nearby the randomly selected points is to physically go to the point and search for them. Any other approach involves the assumption that there is a source of information which lists all villages. For instance, if good maps exist, the random points could be plotted on the map, a circle drawn around each point, and any villages lying within the circle identified. This approach assumes that all villages are marked on the map. If the map is good and contains all villages, then it can be used to create a traditional sampling frame, listing all the villages on the map, and random geographic coordinate sampling shouldn't be used. If the map is not good, and is missing some villages, using the map to select villages will result in an unrepresentative sample.

### Selection radius and stratification

The size of the selection radius, (the area around each point to search for herds or villages) is an important consideration. If a very small selection radius is used (e.g. 500 metres), then most of the points will fall more than 500 metres from a village, and so many points will need to be searched before finding the required number of villages. However, searching the small area around the point is relatively simple. On the other hand, if a large selection radius is used (e.g. 5 km), there is a large area to search for villages, and a danger of missing some. However, many more points will have villages near them, and therefore fewer will need to be searched. The choice of

an appropriate selection radius is a balance between these two considerations. In general, however, the smaller the selection radius, the more confident we can be that no villages or herds have been missed.



The selection radius is also determined by the spread of villages or herds in the study area. If the area is densely populated and villages lie close together, a small selection radius will be adequate. If they are very sparse, then a larger selection radius should be used.

Often, this will take a bit of guesswork. As a rule of thumb, a selection radius of between 1 and 4 kilometres is usually appropriate. The aim is to have only one or two villages within the selection radius of most of the points, and to have only a few points with no villages in the selection radius. There is a practical method for choosing the best selection radius while examining points in the field. It is described under Field Procedures on page 74.

Because the spread of herds or villages may vary from one area to another, it is often convenient to choose a different selection radius for different parts of the study area. This can be done by using stratification by area (described on page 45). The study area is divided into several distinct smaller areas, such as districts. Then in each district, a separate selection radius can be used depending on the spread of villages in that area. Stratification has the added advantage of decreasing the amount of travel involved during field work. When using stratification, the field work is conducted in the same way as described below, but done separately for each stratum.

### Screening using remote sensing

RGCS is used then there is no reliable source of information on where the villages are. That does not necessarily mean that there is no information on where the villages are *not*. Screening is a process of checking the randomly selected points to eliminate those which are extremely unlikely to have any nearby villages. By excluding some of the points with no villages nearby, the amount of time spent in the field searching for non-existent villages can be dramatically reduced.

Some sources of information may indicate areas of land that are unsuitable for villages, or where villages are very unlikely to be. Wide areas of dense forest, or very steep mountainous land are examples of the type of areas that could be excluded through screening.

**Example:** Aerial photographs are used to screen a set of random points. It is not possible to reliably distinguish villages from the photographs, but it is easy to distinguish agricultural land from forest. In the local rice-based agriculture system, it is reasonably assumed that any village will have some agricultural fields nearby. Any random points falling in the middle of dense forest with no nearby agricultural land can therefore be excluded.

#### Remote sensing

There are two possible sources of information that could be used for screening: aerial photography and satellite imagery. Both represent a picture of the surface of the earth. These are examples of *remote sensing*, in which information is obtained about an area (from an aeroplane or satellite) without needing to visit the area. In order to be useful, the remotely sensed information should be:

#### Requirements for remotely sensed data

- Up to date. Old images are of little value because the situation may have changed. If aerial photographs are used to identify forested areas with no sign of agriculture, photographs that are several years old could be misleading, as forest is being cut down at a rapid pace in many parts of the world. Preferably, the images should be no more than a few months old, but in some cases, older images may still be useful.
- Of a suitable scale. More detailed images are harder to obtain, but images with less detail (small scale) may not be good enough to distinguish areas that are unlikely to have a village. Aerial photographs are usually more than adequate for this purpose. There are two readily available types of satellite images, SPOT and Landsat. SPOT images have a *resolution* of 10 or 20 metres, and the resolution of Landsat images ranges from 30 to 80 metres. SPOT images are generally suitable for screening, but Landsat images may not be detailed enough.

**Resolution** is a measure of the size of the smallest feature that can be detected in an image

#### Georeferenced images

- Georeferenced. A photograph of the surface of the earth is of no use unless you know what area the photograph represents. To use remotely sensed data, it is necessary to plot the randomly selected points exactly in their correct location on the images and examine the area within the selection radius. A georeferenced image is one that has information that allows its location on the surface of the earth to be precisely established. All satellite images are georeferenced, and some aerial photographs are as well.
- Inexpensive. Satellite images are often very expensive, costing several thousand US dollars to cover a relatively small area. The cost of having aerial photographs taken of an area is also prohibitively high. The reason for using this type of data is to save costs and field time. If the data cost more than the cost of the field work, there is no point in using them. The only practical way that remotely sensed data can be used for screening is if existing images, originally obtained for another purpose, can be used at minimal expense. Fortunately, the use of this type of data is rapidly increasing in developing countries. In many cases, government departments (such as land resources, or forestry) or research organisations (such as universities) may already have suitable data. These might be obtainable at a very reasonable cost.

If suitable satellite images or aerial photographs are available at reasonable cost, they can be used for screening, either manually, or with the assistance of a GIS. If screening manually, the selected random coordinates are plotted on the image, and

a circle defined by the selection radius drawn around the point. The area inside the circle (and just outside, to account for any distortions, or errors in plotting or georeferencing) is examined for evidence of human activity. Only if it is very unlikely that any village or herd could be located in the area can the point be discarded. If there is any doubt, it should be visited by the field survey team to check.

Plotting and screening the points can be assisted by the use of a GIS if one is available, employing these steps:



- Step 1:** Convert the images to digital form for displaying on the computer. Satellite images are usually available as digital files, which can be imported directly into ArcView. Aerial photographs need to be scanned, and then georeferenced by creating a header file. This file specifies the pixel size and the coordinates of the top left corner of the image. See the ArcView documentation for more information on importing images, and header file format.
- Step 2:** Once random points have been selected and drawn using the **RGCS ArcView** program (described on page 69), the image file can be loaded into the view and displayed. Click the **Add Theme** button, and change the theme type to an image theme. (If your image is in .jpg format, you will need to first load the JPEG reader extension to read the image). Select the file containing the image.
- Step 3:** You can now rearrange the themes so that the random points and surrounding circles are displayed on top of the image. Each point can then be examined individually by zooming in on the circle and examining the area inside.

#### Field procedures

When a list of random points is ready for examination, a field team needs to visit the points and identify nearby villages for selection. This section describes the field activities in detail. A few issues are worthy of consideration when identifying sample villages.

- Random points should be used in the order in which they were selected. The random points created by the computer programs each have a Point ID number, indicating the order in which they were selected. Because many random points may have no villages within the selection radius, more random points were selected than the sample size required, and some of these points will not be used. While it may be more convenient to first visit some of the random points with a high ID number, all the points with a low ID number should be used first. If you choose some of the random points to use (because they are convenient) and ignore others, the selection is no longer random.
- It is important to identify and record *all* the villages or herds that lie within the selection radius of each random point used. Even though only one village is required, that village can be thought of as representing all the nearby villages. The total number of nearby villages within the selection radius is used to weight the results from the selected village. This indicates how many similar villages the selected village is representing. For this reason, all villages within the selection radius must be identified.
- For each village or herd visited, it is necessary to determine the distance from the random point. For small herds this may be straight-forward. However, some villages are large, and may extend for several kilometres. If part of the

village is within the selection radius and part is outside, there must be a rule to determine whether to count the village or not. This is done by determining a single unique point in the village that is used to determine its location. The choice of point will depend on the culture and situation. Some possibilities are a central market, intersection, school or place of worship, but these are only useful if every village has one, and no village has more than one. A good choice in many cases is the house or office of the chief or head of the village. There is normally only one head, and most residents know where it is, so it is easy to locate.

With these issues in mind, the procedure for identifying villages is as follows:

**Step 1:** Plan a route. It is useful to first plot all the random points on a road map, along with the point IDs and plan a route that will take in as many of the points as possible. The route should include all points numbered up to the sample size and as many of the other points as possible, with highest priority given to the lower numbered points.

**Example:** If the sample size is 5, points 1 to 5 must be included on the route, and priority should be given to visiting points 6 to 10. It is initially less important to visit points 11 to 15, as these will only be needed if there are fewer than 5 villages near the first 10 points.

**Step 2:** Prepare recording sheets. Two sheets are required, as shown below. Sheet 1 is used to record **villages** which have been identified near one of the points. Sheet 2 is to keep track of which **points** have been visited, and how many villages are nearby. Full copies of both sheets are included in Appendix D.

**Sheet 1:** Village sheet for recording the name of each village visited, and its distance from any nearby points:

Sheet 1: Villages				
Village Name	Point Number	Distance from point	Rank	Selected

**Sheet 2:** Point sheet for recording which points have nearby villages

Sheet 2: Points		
Point Number	Visited (Yes/ No)	Number of villages
Point 1		
Point 2		
Point 3		

- Step 3:** Prepare the GPS unit. Many units have the ability to store a number of geographic points or “landmarks”. If possible, the coordinates of the random points can be entered into the GPS, and identified with their ID number.
- Step 4:** Visit the villages in each selection radius. Visit the points in the order dictated by the route. It is not necessary to visit the points in the order they were selected, as long as all the low-numbered points are visited at some time. When visiting points, it is not necessary to identify the exact location of each point, as it will often be impossible to access. It is adequate to identify and visit all the villages that lie near each point.
- Step 5:** Visit points. For each point that lies on or near the chosen route, mark the point as visited when the GPS indicates that you are closer than the chosen selection radius (write “yes” in the Visited column next to the point ID on Sheet 2 - Points) . The selection radius used should be larger than you think will be necessary, say 3 or 4 kilometres. After the points have been visited, the selection radius can be changed to a smaller value, as described below. Start searching for villages near that point.
- Step 6:** Visit villages. Write the name of each village that you visit within the selection radius of each point on Sheet 1 - Villages.
- Step 7:** Locate the unique point identifying the location of the village (e.g. the village head’s house). While at this location, note the distance to the nearest point, as reported by the GPS, and record the point ID and distance next to the village name on Sheet 1 - Villages.
- Step 8:** While still at the same location, check for any other nearby points. It is possible for one village to fall within the selection radius of two points. If there are any other nearby points, record the village name, point ID, and distance on a new line on Sheet 1 - Villages.
- Step 9:** Search for other nearby villages. This is best done by asking residents for directions and approximate distances to any other nearby villages. It is also advisable to explore roads for other villages. When all nearby villages have been identified and recorded, write the total number of nearby villages on Sheet 2 - Points, on the line for that point.
- Step 10:** If no villages are found nearby a point, mark the point as visited on Sheet 2 - Points, and record the number of villages as 0.
- Step 11:** Continue until at least all the points with low ID numbers have been visited.
- Step 12:** Check to see if enough points have been visited. Using Sheet 2 - Points, work down the list from the top to the first point that hasn’t

yet been visited. Ignore later points that have been visited. Count the number of points with one or more villages, up to the first (lowest numbered) un-visited point. If the number (of sequentially-visited points with nearby villages) counted is greater than or equal to the required sample size, then enough points have been visited. Otherwise, continue visiting points.

**Example:** The table below shows how Sheet 2 - Points might look after 7 points have been visited. To check to see if enough points have been visited, points 1 to 5 would be counted, but points 7 and 8 would be ignored, because point 6 has not yet been visited. The total number of points with at least one village within their selection radius is 3 (points 1, 3 and 4), as there were no villages near points 2 and 5. If the required sample size is 5, then more points would need to be visited. The next point to visit should be point number 6.

Sheet 2: Points		
Point Number	Visited (Yes/ No)	Number of villages
1	Yes	1
2	Yes	0
3	Yes	2
4	Yes	1
5	Yes	0
6		
7	Yes	1
8	Yes	2
9		
10		

**Step 13:** Visit more points. Even if enough points have already been visited, if it is practical to continue visiting more points (time permitting, or they are conveniently located) visit more points. This may enable a smaller final selection radius to be used. Otherwise, the information collected can now be used to determine which villages to include in the sample.

#### Selecting sample villages

**Step 1:** Identify eligible points. On sheet 2 - Points, find the first point that was not visited. All information from points later than this (with a larger Point Number) should be discarded. Of the remaining points (with a smaller Point Number), identify all which have at least one village within their selection radius.

**Example:** After visiting several more points, the data collection sheets 1 - Villages and 2 - Points look like those shown below. All data after point 9 would be discarded (points 10 and 12), and only points 1 to 8 considered. Points 1, 3, 4, 6, 7 and 8 have at least 1 village in their selection radius.

Sheet 2: Points		
Point Number	Visited (Yes/ No)	Number of villages
1	Yes	1
2	Yes	0
3	Yes	2
4	Yes	1
5	Yes	0
6	Yes	1
7	Yes	1
8	Yes	2
9		
10	Yes	2
11		
12	Yes	1

Sheet 1: Villages				
Village Name	Point Number	Distance from point	Rank	Selected
Ban O	7	2.13		
Sisavat	1	1.82		
Ban Pak Ngeum	8	0.41		
Ban Hai	3	0.83		
Tong Kham Khong	4	1.57		
Sithay Noy	3	2.55		
Nong Bone	8	2.31		
Talath	6	1.39		

**Step 2:** Mark closest villages. On Sheet 1 - Villages, put a cross in the Point Number column next to the closest village to each of these points.

**Example:** Using our data, there are six villages (marked with a cross). Village Sithay Noy and Nong Bone are not marked as there are other villages closer to their points.

Sheet 1: Villages				
Village Name	Point Number	Distance from point	Rank	Selected
Ban O	7 X	2.13		
Sisavat	1 X	1.82		
Ban Pak Ngeum	8 X	0.41		
Ban Hai	3 X	0.83		
Tong Kham Khong	4 X	1.57		
Sithay Noy	3	2.55		
Nong Bone	8	1.31		
Talath	6 X	1.39		

**Step 3:** Rank the villages. Rank the villages marked by a cross according to their distance from the closest point by writing a number in the Rank column. Write 1 in this column for the village closest to any of the marked points. Write 2 for the next closest and so on.

Sheet 1: Villages				
Village Name	Point Number	Distance from point	Rank	Selected
Ban O	7 X	2.13	6	
Sisavat	1 X	1.82	5	
Ban Pak Ngeum	8 X	0.41	1	
Ban Hai	3 X	0.83	2	
Tong Kham Khong	4 X	1.57	4	
Sithay Noy	3	2.55		
Nong Bone	8	1.31		
Talath	6 X	1.39	3	

**Step 4:** Determine the optimal selection radius. Find the village with a rank equal to the sample size. The distance for this village is the optimal selection radius. This is the smallest radius that may be used and still result in the selection of the number of villages required.

**Example:** The marked villages have been ranked from 1 to 6. The sample size required is 5 so the village with rank 5 (Sisavat) is used to calculate the selection radius of 1.82 km.

**Step 5:** Identify all villages within the selection radius. All villages nearby the points under consideration are checked. Those with a distance less than or equal to the selection radius are marked by a tick in the Distance column.

**Example:** The villages with a distance less than or equal to 1.82 km have been marked with a tick in the distance column. Note that even though Nong Bone is not the closest village to point 8, it is still within the selection radius, and therefore marked.

Sheet 1: Villages				
Village Name	Point Number	Distance from point	Rank	Selected
Ban O	7 X	2.13	6	
Sisavat	1 X	1.82 ✓	5	
Ban Pak Ngeum	8 X	0.41 ✓	1	
Ban Hai	3 X	0.83 ✓	2	
Tong Kham Khong	4 X	1.57 ✓	4	
Sithay Noy	3	2.55		
Nong Bone	8	1.31 ✓		
Talath	6 X	1.39 ✓	3	

**Step 6:** Select the villages. If there is only one selected village around a particular point, then that village is included in the sample. If there is more than one, one of the villages must be randomly selected. Use a random number to pick one of the villages (flip a coin if there are only 2, or roll dice if there are less than 6 – otherwise use a computer or random number table).

**Example:** The selected villages have been marked on the Sheet 1 - Villages. There are two villages within the selection radius around point number 8, Nong Bone and Ban Pak Ngeum. A coin was tossed to determine which village should be included.

Sheet 1: Villages				
Village Name	Point Number	Distance from point	Rank	Selected
Ban O	7 X	2.13	6	
Sisavat	1 X	1.82 ✓	5	Yes 1
Ban Pak Ngeum	8 X	0.41 ✓	1	
Ban Hai	3 X	0.83 ✓	2	Yes 1
Tong Kham Khong	4 X	1.57 ✓	4	Yes 1
Sithay Noy	3	2.55		
Nong Bone	8	1.31 ✓		Yes 2
Talath	6 X	1.39 ✓	3	Yes 1

**Step 7:** Calculate weights. Villages are assigned weights equal to the number of villages within the sampling radius of the point. If only one village falls within the sampling radius, then the weight is equal to one. If three villages are within the radius, then the randomly selected village gets a weight of three. These weights are used to avoid bias in the calculation of estimates.

The procedure described here is used to select sample villages before the start of the survey data collection. This is most practical when distances are relatively short, and villages need to be notified in advance of a visit, to organise a village meeting for instance. Sometimes it may be more practical to conduct the field data collection at the same time as the village is selected. This is particularly so when travel to the village is difficult. Unfortunately, when surveying villages at the same time as selecting them, it is necessary to visit each random point in the order in which they were listed, and decide on the selection radius before the start of field work. Both these factors are likely to increase the amount of travel involved in the survey work.

# 4

## Principles of Data and Specimen Collection

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Sources and quality of information

Recording data

Animal restraint

Specimen collection and processing

Transporting and processing blood

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



## Sources and quality of information

There are various sources of information about livestock diseases and different ways of collecting this information. In general, the quick and easy ways provide data that are less complete or less reliable. If good quality data are needed, collecting them usually requires more time, effort and expense.

### Existing records

The easiest way to gather data is to use data that somebody else has already collected. When considering disease information, this is known as passive surveillance, discussed in Chapter 1. Records that may already exist include outbreak reports, laboratory submissions, and livestock population.

The problem with using existing records is that the reason the data were collected usually differ from our purpose. For instance, livestock population data may have been collected from the villages for the purpose of reporting total numbers within a province. It is therefore only available as provincial totals, with no individual village totals. To be useful for surveys, individual village populations are needed. When we use existing records, we have no control over how they were collected or how complete the information is. Active surveillance is used to overcome these problems.

### Interview data

Better quality and more relevant information can be obtained by collecting it specially to answer particular questions. A relatively quick and simple way of collecting information is to ask people who know. The quality of information collected through interviews depends both on *who* is asked, and *how* the question is asked. For livestock disease surveys, there are various people that know about the disease situation. Livestock owners know the most about their animals, but may not have very good knowledge or training about livestock diseases. Village veterinary workers who have received basic veterinary training may be more knowledgeable about diseases than the owners, but may not have specific details about individual animals. Local veterinary officers usually have much better technical knowledge of disease and production, and an overview of disease problems in their area. Provincial officers are likely to be even better trained, more experienced, and have a good understanding of the disease situation in their province. However, at each level up the hierarchy, even though training and technical knowledge increases, contact with the actual animals decreases.

**Example:** A survey is planned to find out which diseases cause the greatest impact in the village smallholder system. Funds and time are limited, so a written questionnaire is sent from the central office to provincial veterinary offices throughout the country. The questionnaire is addressed to the provincial veterinarian, and asks for information on which are the most important diseases in the villages. There are only 26 provinces, so the survey is easy to prepare, and all the responses are received back at the central office within 2 weeks. When the data are analysed, the results closely match the government's published list of priority areas for disease control.

The result is not surprising, and may not be particularly useful. The provincial officers know the government's disease priorities, as do their district staff. Most of the disease reports received, or calls for assistance will be related to one of these priority areas, because it is known that these are important areas for the government. Unfortunately, this may not be the same as the most important diseases from the livestock owners' point of view. The survey failed to ask the people who really had the information required. The same survey could have been conducted by asking district staff, village veterinary staff, or the livestock owners themselves. Each of these options would be more difficult, involving progressively more people and expense, but the quality of the information would be better.

In some situations, particular groups of people are the best source of information, as they are the ones that hold the key information. For instance, if evaluating funding needs for local veterinary offices, the staff of those offices are the ones who know the work and expenses of the office best, just as livestock owners know about the husbandry of their animals best.

Collecting information from people has a few added complications arising from human nature. When asked a question, a person may give the wrong answer for a variety of reasons:

- They have forgotten about something that happened some time ago.
- They may not know the answer, but don't want to admit it.
- They may lie on purpose, because they are afraid about how the information will be used.
- They may not understand the question.
- The interviewer may not understand the answer.

To overcome these problems, interviewers should be aware of what problems can occur, and use techniques to avoid them. One such technique is the use of a village interview, at which many livestock owners are asked about their animals at the same time. Village interviews are discussed in detail in Chapter 5.

## **Examining animals**

Although livestock owners know their animals better than anybody else, they may not be skilled at identifying particular diseases. When we are interested in collecting reliable information about animal diseases, it is sometimes better to consult the animal directly, rather than the owner. The advantage of directly examining animals during a disease survey is that much better quality data can be collected. There is no longer the need to depend on the farmer to make a diagnosis.

A disease survey can be conducted by just a few survey teams over a relatively short period of time. It is therefore possible to have survey staff who are well trained and highly skilled in diagnosing diseases (perhaps more so than the district or provincial staff).

## **Collecting specimens**

It is not possible to diagnose all diseases simply by carrying out a clinical examination of an animal. Many diseases can be diagnosed only with the use of laboratory tests.

Often, surveys are not concerned with clinical disease, but may be aimed at measuring the prevalence of subclinical disease or evidence of past exposure to a disease or vaccination. In these cases, laboratory testing of specimens (such as blood

or faeces) collected from animals is necessary. The quality of information gathered through the collection of specimens and the use of laboratory tests is usually better than is available through any of the other means. Frequently, laboratory testing of specimens is the only way to provide a definitive diagnosis of the disease. On the other hand, surveys using specimen collection and laboratory tests are often expensive and time consuming. When planning a survey, it is important to evaluate what resources are available, and how good the quality of the information needs to be. The four different approaches to collecting data are summarised below.

	Expense	Speed	Up to date?	Accuracy
Existing data	Very cheap	Very fast	Often out of date	Not good
Interview data	Moderately expensive	Fast	Usually up to date	OK
Examining animals	More expensive	Slower	Very up to date	Good
Collecting specimens	Most expensive	Slowest	Very up to date	Best

## Recording data

Collecting information during survey field work is often difficult and exhausting work. There is nothing worse than to find at that end of the field work that the information wasn't recorded properly, or that sheets have been lost, and all that hard work is wasted. Before collecting any information, it is important to plan how the information will be recorded, and to have a well-organised system for keeping the records safe.

When planning the survey, decide on what sorts of information will be collected and how it is to be recorded. Based on this, design sheets with places to write down all the information. The design of the data recording sheets should make it easy to record the data in the field, and easy to analyse the data afterwards. A series of example data recording sheets for the types of surveys described in this book is shown in Appendix D. These can be copied and used directly, or else modified for your specific needs. It is useful to be aware of a few general guidelines when using data recording sheets:

### Village / Herd codes

- Every sheet should be marked with the source of the information. This may be the name of the village, herd, or household. It is also useful to use a numeric code to identify all the villages or herds in the survey. This prevents problems with misspelling names, and makes it easier to enter the information into a computer for analysis (see Chapter 6 for more on using computers for analysis). Other useful information to include on each sheet is the date, and the survey team (if more than one is used).
- Always write down information as soon as it is collected, and never trust it to memory. For instance, during a village interview, lots of information can be collected in a short time. The person conducting the interview may be too busy talking with the livestock owners to write down all their answers. In this situation, one member of the survey team should lead the interview, and another should be responsible for recording all the information.

- Make sure that each member of the survey team knows their responsibility. One person should be made responsible for writing down the information while others may be responsible for collecting it.
- Write the information clearly and carefully. Use a clipboard as a surface for writing on to make it easier. If the information is hard to read when being analysed, mistakes may be made.
- Keep the sheets safe and dry. At the end of each day, all completed sheets should be stored away in a folder in a safe place, or sent directly for data entry.
- Don't make a 'neat copy' of the data. Every time that information is copied by hand, there is the risk of making a mistake and writing down the wrong information. If you want a copy of the information (which is a good idea, in case the original gets lost), make a photocopy. Make sure that the information is clear enough to read the first time.
- When collecting blood or other specimens, label each tube carefully with a permanent marker that won't come off if the tube gets wet. Use an identification number on the tube, and record the same number on the data sheets, so that it is clear where the blood came from.

## Animal restraint

When either examining animals or collecting specimens, the animals must first be restrained. Veterinary staff are often experts at restraining animals, and there are many useful approaches and tricks. Livestock owners often know little about effective restraint, as they rarely have to examine their animals closely, or collect specimens. Intensive farms are likely to have reasonably good facilities for handling and restraining animals, but in villages there are often no facilities at all.

Race or crush

One approach for large animals is to construct some sort of race or crush, where animals can be kept still. This may be practical when regular visits are made to a particular village. In the normal situation, a disease survey examines animals from a number of different villages, and may only examine them once. As the survey team visits for only one day, the time and effort required to build a crush is not warranted. This section describes a few quick and simple restraining techniques that have been used for disease surveys.

### Cattle and buffalo

The amount of restraint needed depends on what is being done to the animal, the size and temperament of the beast, and the experience of the operators. Calves may need little restraint and can be handled similarly to goats. Very quiet animals may also be examined with little restraint. However, for painful procedures, like collecting blood, it is usually necessary to fully restrain the animal. Here are a few techniques that may be used on their own, or in combination. There are many other ways to restrain an animal, and the best one to use depends on the animal, the situation, and the operator's experience.

Nose halter

**Permanent halters**

In some countries, cattle and buffalo wear a permanent halter to assist in everyday handling. This may be a simple rope halter, a custom made halter with straps, or a *nose halter*. The nose halter passes through a hole in the nose and around the back of the head. By attaching a rope to the halter, or holding it directly, it is possible to lead animals. However, use of the halter is rarely enough for blood collection, as the animal can swing around and is still able to pull away. Permanent halters are often made of relatively weak material, and can break easily if the animal jerks because it is hurt or frightened.

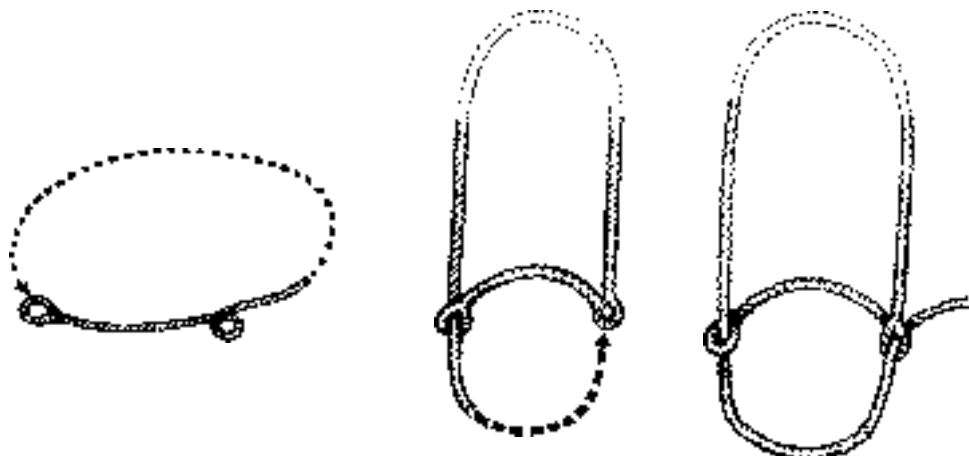


**Removable halters**

A removable halter made of strong rope is very useful for controlling cattle and buffalo, or for leading them into a crush. The halter can be easily made from a length of rope, and should have at least 3 metres of lead rope attached. To make a halter:

Making a rope halter

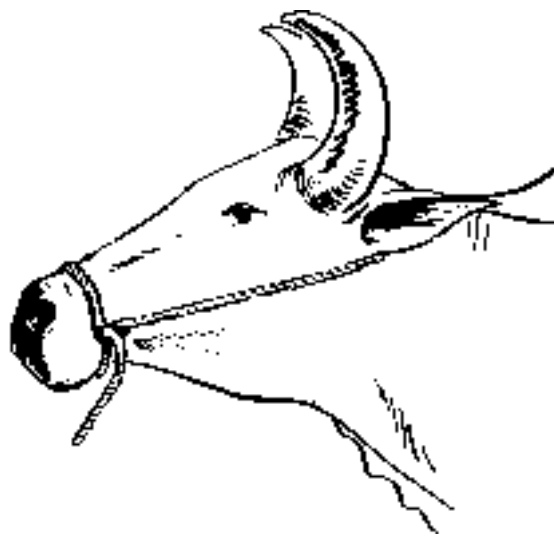
- Step 1:** Take a length of 4 to 6 metres of strong (about 10 mm diameter) rope.
- Step 2:** Make a small loop in one end.
- Step 3:** Make another small loop in the rope about 20 cm from the first.
- Step 4:** Pass the long end of the rope through the first (end) loop, and then through the second loop.



To put the halter on:

Putting on a halter

- Step 1:** Arrange and hold the halter so that the second small loop will be on the left side of the animal's head. The lead rope should pass in the second small loop, under the chin of the animal, through the first loop, then back over the head behind the ears.
- Step 2:** When the halter is ready, approach the animal quietly, and place the halter over the head behind the ears.
- Step 3:** Pass the loop under the chin, and tighten.



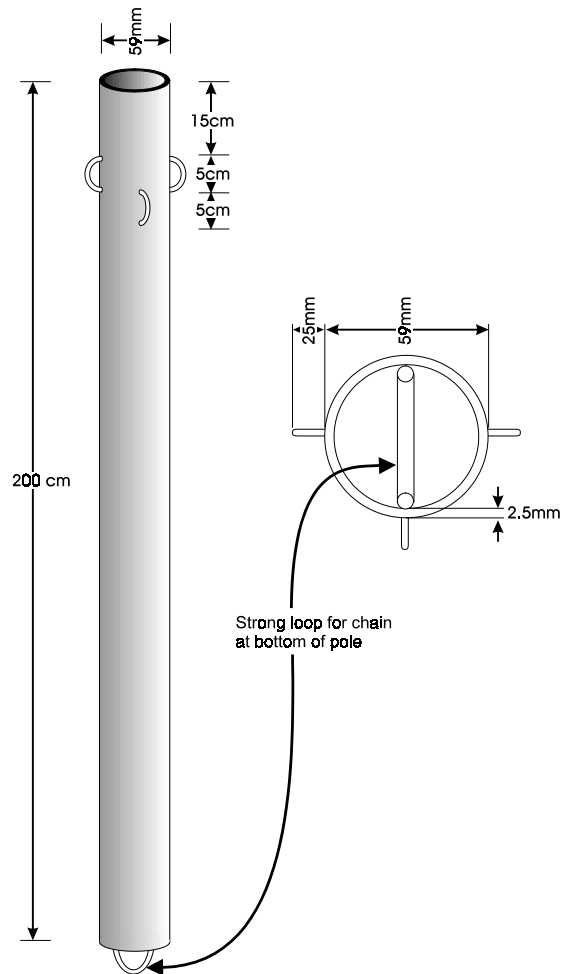
It is easy to put a halter on the wrong way, but it won't work unless put on correctly. Make sure that there are no twists and it is comfortable on the animal. When it has been put on properly, you can pull on a halter very hard without any risk of hurting the animal.

A removable halter is very good for tying up an animal, or leading it somewhere. However, even when tied up, the animal is still able to swing its body around, and shake its head. Usually more restraint is needed before blood can be collected.

### **Bleeding pole**

One very useful piece of equipment for restraining cattle and buffaloes when there are no handling facilities is the *bleeding pole*. This is a 2 metre long metal pole (made out of steel water pipe) which is placed against a tree or other solid support. It is used to capture the neck of the animal, like a head-bail.

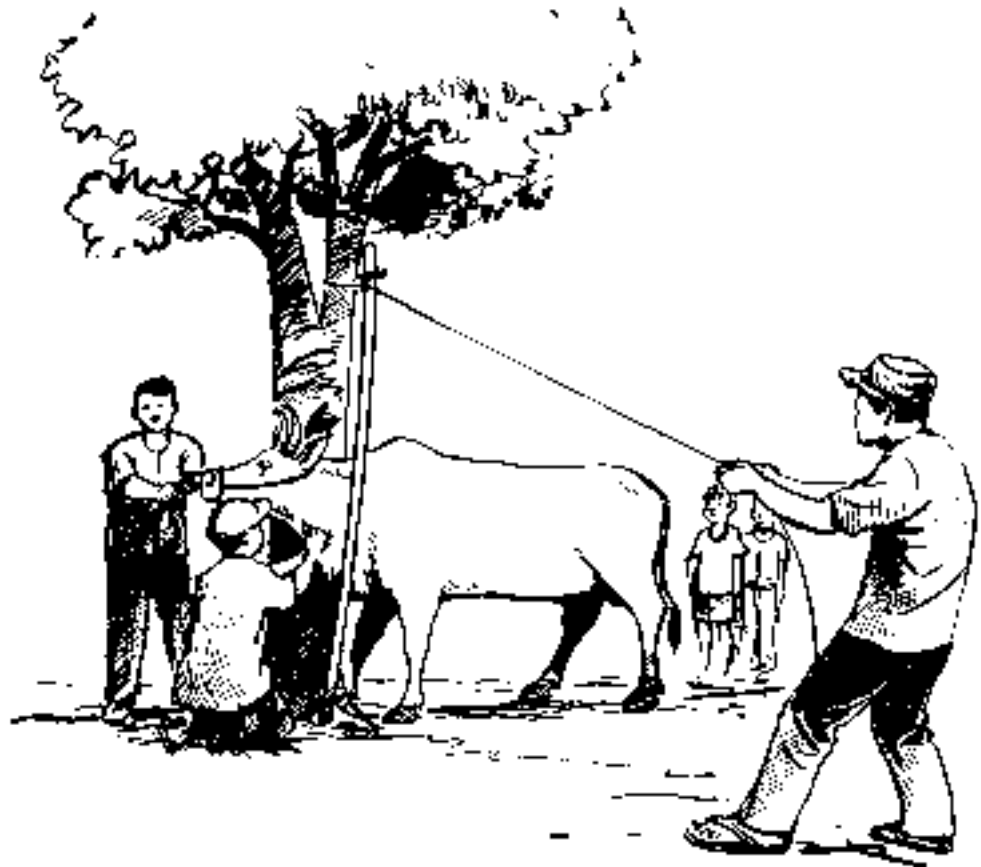
A bleeding pole can be made by welding 4 loops of metal onto a 2 metre length of steel pipe, as shown in the figure below. To use the pole, you need a length (about 1.5 metres) of strong chain, fitted with a hook at one end, and about 4 metres of strong (10 mm) rope.



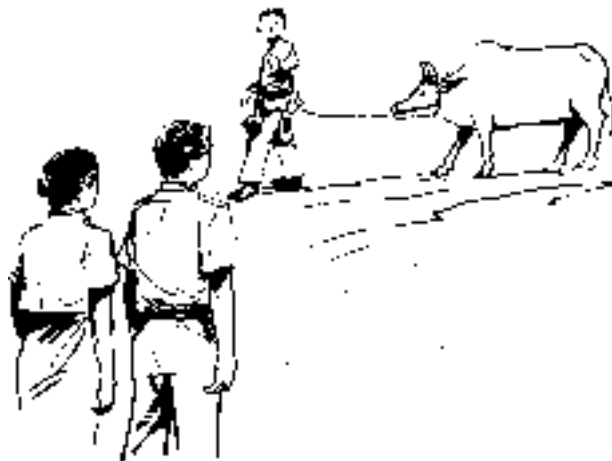
To use the bleeding pole:

- Step 1:** The bottom end of the pole is rested on the ground at the base of a sturdy tree.
- Step 2:** The chain is passed through the bottom loop, around the tree, and fastened to itself with the hook. This stops the bottom of the pole from moving away from the tree, and allows it to hinge outwards
- Step 3:** A rope is fastened to the pole, passing through the lower of the three top loops (on the side away from the tree).
- Step 4:** The rope is then passed around the tree, and back through either of the two top loops on the side of the pole. The end rope is then held well away from the tree and the pole. When the rope is pulled, the pole is held close against the tree.
- Step 5:** The rope is loosened, and the pole is leant away from the tree at an angle of 45° or more.
- Step 6:** A halter or lead rope is used to bring the animal towards the tree. The animal's head is passed between the tree and the pole. For buffaloes or other animals with wide horns, the head may need to be twisted to fit between the tree and the pole.

**Step 7:** When the animal's head is through, the rope is pulled to tighten the pole against the tree, and capture the animal. To secure, the rope can be passed around the tree again, or looped around the top of the pole.



It is sometimes difficult to pull the animal's head between the tree and the pole, and sometimes the animal may leap through the gap. It is often easier if the animal's owner brings the animal after the pole has been set up, and that the pole is closed quietly against the tree so as not to frighten the animal. Minimising the amount of noise and activity, and keeping unfamiliar people away from the animal will keep it calmer and make the job easier.



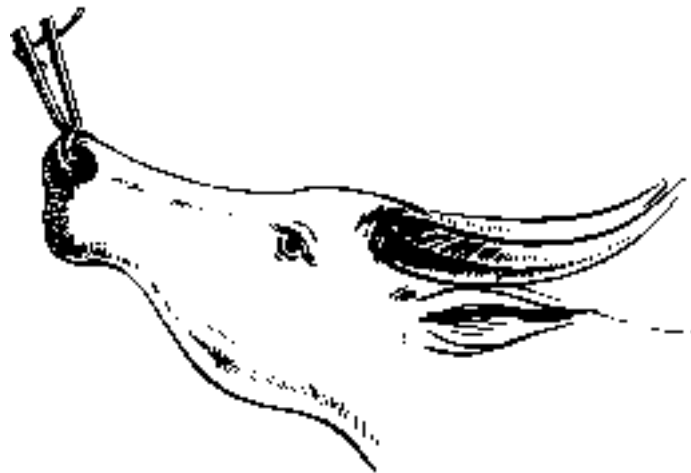
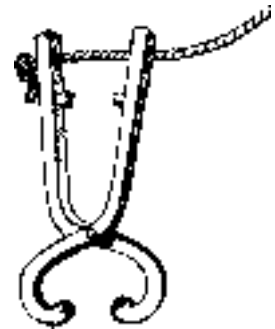
If the pole is kept tight, an animal cannot escape. Using the bleeding pole is usually enough to restrain an animal for examination. To take blood, from the neck, it may be necessary to restrain the head to stop it moving, and also to have somebody pull the tail to one side (the side the tree is on), to stop the back end of the animal swinging around.

The pole has the advantage that it is strong, reasonably light (it can be carried by one person for a reasonably long distance), cheap to produce, and easy to transport. If no specially made bleeding pole is available, a length of strong wood can be used in the same way. Make sure that it is strong enough before using it.

**Nose grip**

To keep an animal very still, and prevent movement of the head, *nose grips* can be used. Nose grips are like a large pair of pliers with round blunt ends that don't quite close at the end. They are placed inside the animal's nose, and tightened. They can then be used to hold the head in any position necessary.

Nose grips provide very good restraint and control over the animal, and should be used whenever necessary. A small proportion of animals don't like them, and become more agitated, so take them off if they aren't helping.



**Pigs**

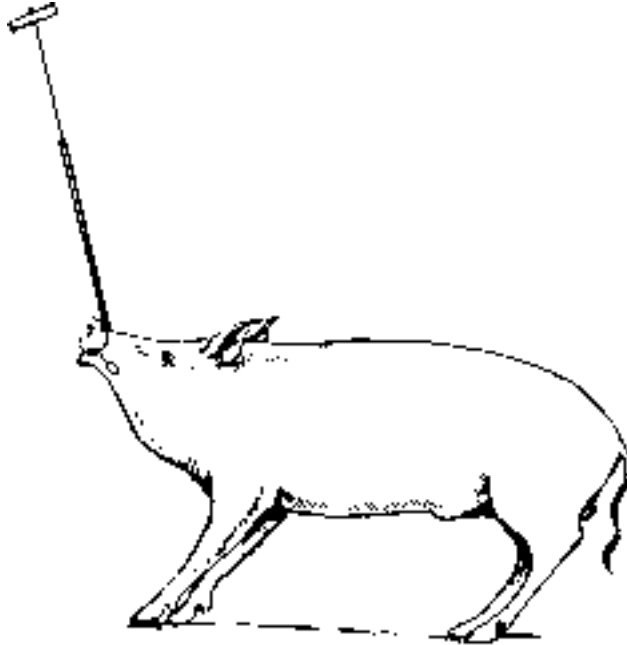
Pig snare

Piglets and small pigs can be easily managed by simply holding them. Larger pigs can be difficult to restrain, and may need several people. One useful piece of equipment is the *pig snare*.

This consists of a loop of steel cable which is passed through a tube, and attached to a handle. To use a pig snare, the loop is loosened, and slipped inside the mouth of the pig. Pigs are inquisitive and will usually taste and bite the loop if it is put near them. The loop should be behind the canine teeth, and is then tightened over the top of the snout. The handle is held up or tied. Be sure that the cable is thick enough not to cut into the mouth of the pig.



When captured in this way, pigs will pull back very hard, and scream loudly. If working near them (for instance collecting blood) you should wear ear plugs to protect your hearing. Even though only the snout is captured, the pig will usually stay very still. If a pig snare is not available, a simple length of rope with a sliding loop in one end can be used in a similar way.



## Horses

Horses are usually more used to being handled by people than other animals, so examination is simpler.

### Halter

A halter, the same as that used for cattle is usually enough to restrain a horse for examination or blood collection. If better restraint is needed, a twitch can be used.

### Twitch

A twitch is a way of grabbing and twisting a piece of skin. This causes discomfort, and the horse will usually stand very still. A simple way to use a twitch is to grab some skin on the neck and twist it, or to grab one ear and twist it. A more effective method is to use a nose twitch. A nose twitch is a wooden stick with a small loop of rope passed through a hole at one end. To use it, place the loop over the fingers of the left hand, then grab the muzzle and upper lip with the left hand and slip the loop off the fingers onto the muzzle. With the right hand, twist the stick to tighten the rope. A horse will usually stand very quietly with a nose twitch.



Be careful not to use it for too long or too tight, as you could hurt the delicate skin of the muzzle.



### Sheep

Sheep are relatively easy to restrain, but may be hard to capture. When captured, pick the sheep up, and rest it on its rump, with its back leaning against your legs. The sheep will sit quietly. For blood collection, the head can be held back behind one arm to expose the neck.



### Goats

Goats are also small enough to restrain relatively easily. They can often be kept still just by putting an arm in front of their chest. For any painful procedures, they can be held on their side. To place a goat on the ground, first hold the goat with its side against your legs. Put one hand over the back, down the far side, and grab a fold of skin under the belly. Put the other hand under the neck. Pushing with your knee, lift the goat off its feet and down onto its side on the ground. As you do so, you can hold the head and fold it back onto the body. By keeping the head pressed down, the goat will not be able to stand up. You can lay the head flat on the ground, and use your knee to keep it down. Be careful of the legs when holding goats this way, as they can still kick.



Calves and even medium sized cattle can be held on the ground in the same way.

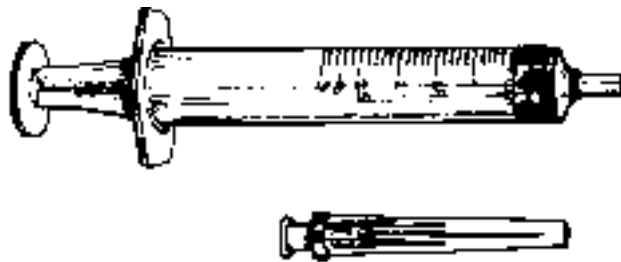
## Specimen collection and processing

A blood sample is the most common type of specimen collected during livestock disease surveys. Blood can be analysed in the laboratory to identify a wide range of antibodies which can be present due to disease or vaccination, as well as antigens from different organisms, and other signs of disease. Compared with other types of samples (such as tissue samples or swabs), it is relatively easy to collect and transport. However, you usually only have a chance to collect blood once during a survey. If the sample is not large enough, or has been destroyed during transport to the lab, then the information in the blood is lost, and the effort selecting animals and collecting the blood has been wasted. It is therefore important to collect enough blood, and handle it carefully while transporting it to the laboratory.

Specimens other than blood can be collected. Faeces is examined to detect the eggs of intestinal parasites, and is easy to collect and transport. Other specialised samples, such as nasal swabs, oesophageal scrapings or tissue biopsies require specialised techniques, equipment and transport media. If you intend to collect these specimens, work with the laboratory staff to plan the best approach.

Blood collection with a syringe

There are two main pieces of equipment collecting blood. A normal needle and syringe can be used in any situation, and, with practice, can be used with one hand. When using a syringe, make sure to transfer the blood to a tube for transport as soon as it has been collected. Cap the needle, and remove it before squirting the blood into the blood tube.



Evacuated tube

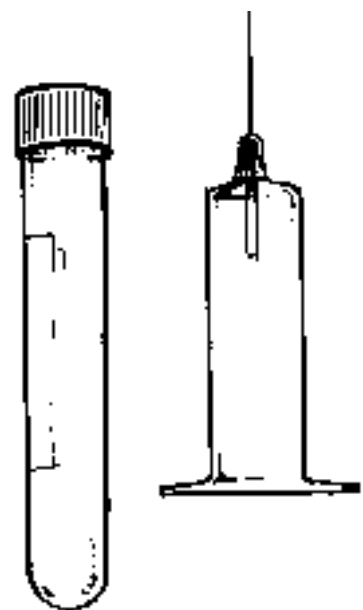
The alternative is to use *evacuated tubes* (or *vacuum tubes*). These are empty glass or plastic tubes, with a rubber stopper. All the air has been removed from the tube so there is a vacuum inside. They are used with a special holder, which is a plastic sleeve fitted with a two-ended needle.

To use an evacuated tube:

**Step 1:** First place a needle in the holder. The short needle fits inside the sleeve, and the long needle is outside. Remove the cap on the short needle, place it in the holder, and screw it in to tighten it.

**Step 2:** Insert the evacuated tube into the holder. If the tube is pushed all the way in, the short needle will puncture the stopper, and air will

enter the tube, making it useless. There is usually a fine line marked near the top of the holder. You can push the tube up to that line



without letting air into the tube. The needle is half-way through the stopper, so the tube won't fall out.

- Step 3:** Remove the cap on the outer needle, and place the needle into the vein of the animal. Even when the needle is in the vein, no blood will flow.
- Step 4:** Push the tube firmly into the holder, so that the inner needle goes all the way through the stopper. The vacuum then sucks the blood into the tube. Hold the tube still until it is almost full. If no blood flows, move the needle slightly to try to find the vein. Don't pull the needle out, or else air will enter the tube and it will be wasted. If you want to pull the needle out and start again, first pull the tube out of the holder, then pull the needle out of the animal. Needles will often become blocked with a small amount of blood when they have been used once, so if you are trying again, it is usually best to put on a fresh needle.
- Step 5:** When the tube is full, pull the tube and needle out of the animal and press the site for a few moments to stop any bleeding.

Evacuated tubes are more expensive than normal syringes, but they are often faster and simpler to use, as you don't need to pull back on the plunger to take blood, and you don't need to transfer the blood from the syringe to another container for transport.

The best place to collect blood differs between species, and there are often several suitable places. Some commonly used sites are described later for the different species. There are however, some general guidelines for all species:

#### Guidelines for blood collection

- Always use a fresh needle for every animal. Reusing the same needle can spread disease between animals, can contaminate the specimen by mixing the blood from two animals, can be frustrating, because the needle will often be blocked with blood, and can be difficult as needles get blunt quickly.
- Always use a new syringe for every animal. Reusing an old syringe can spread disease, and contaminate the sample. Plastic syringes can be cleaned and sterilised, but they quickly become stiff and difficult to use, so re-use is best avoided.
- Restrain the animal properly before getting blood. If the animal moves, it is much harder to collect blood, and may be dangerous to the animal or to you.
- Make sure the skin of the animal is clean before taking a blood sample. Usually, if it appears clean, there is no need to do anything. If there is any dirt on the skin, it can be scraped off and then cleaned with some alcohol on cotton wool or a swab. Let the alcohol dry before inserting the needle. Some animals are more susceptible to infection than others. With horses, always clean the site with alcohol before inserting the needle.
- Blood clots quickly after it has been collected. If the serological test used by the laboratory needs serum, then plain blood tubes should be used to allow it to clot. If plasma, or blood cells are required (e.g. white blood cells for antigen detection tests), then an anticoagulant should be used to prevent clotting. Check with the laboratory which anticoagulant is preferred – common ones include heparin and EDTA. When using an anticoagulant, the blood should be *gently* mixed once it is in the tube. Rock the tube from end to end three or four times to mix. Don't shake the tube, as this will break the cells and ruin the sample.

- Used needles should be disposed of carefully. There are various diseases that can be passed from animals to humans through being pricked by a needle, so be very careful. Never throw needles on the ground, or leave them lying around. A plastic bottle with a narrow neck (like those used to sell water or soft drinks) makes a good container for used needles. When full, it should be burnt in a hot fire to destroy the needles.

## Cattle and buffalo

Jugular vein

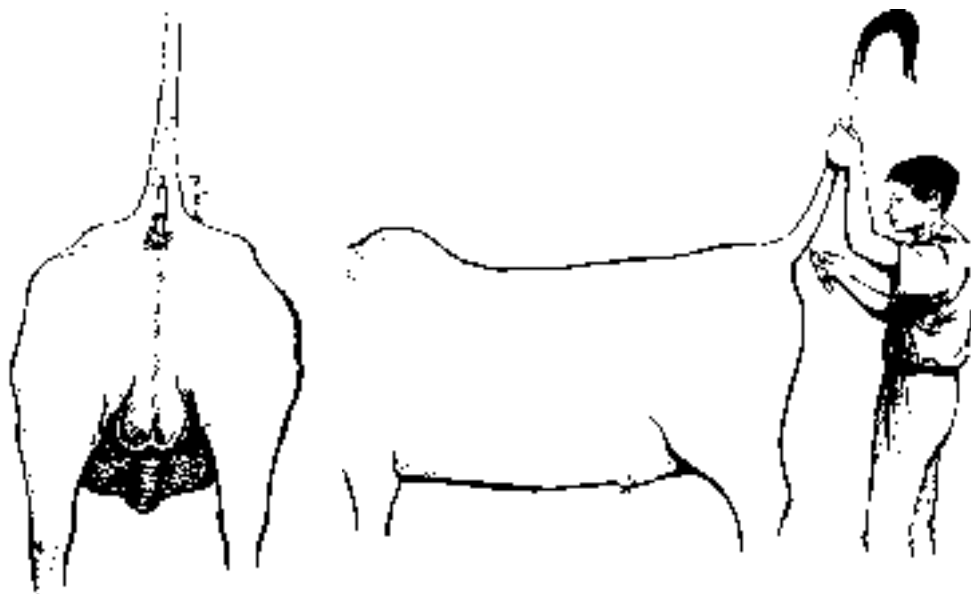
Blood can be collected from two places in cattle – the jugular vein in the neck or the caudal vein on the underneath side of the tail. To collect blood from the jugular vein, the animal should be restrained, and the head held upwards. Use nose grips, and pass the rope over a branch to pull the head upwards. To stop the animal moving its head, it is often useful to have an assistant hold the horns and muzzle firmly. The jugular vein runs in a groove on the lower part of the neck on each side. You may be able to see a pulse in the part closest to the chest. The vein is easier to see when you place your fingers over the vein near the chest and press. This blocks the flow of blood, and the vein fills up. The vein is very large and is usually quite easy to see, and to place a needle in. Use an 16 to 18 gauge 1½ to 2 inch needle. It is possible to collect a lot of blood in this way if necessary. In large cattle, and buffaloes, the skin can be thick, and the vein hard to see. You may need to use a 2 inch needle in these cases.



Caudal (tail) vein

Collecting blood from the tail vein is an alternative method. To do this, the animal must be held so that it is not able to swing its back end from side to side. If a race is available, or some sort of crush with side rails, this is often faster than using the jugular, as the head doesn't have to be restrained. Approach the back end of the animal carefully to avoid being kicked. Grab the tail, and push it straight up – be careful that it is not twisted to one side. When the tail is pushed up firmly, cattle will not kick, and it is quite safe to stand behind them. Clean an area about 4 cm from the

base of the tail exactly in the middle of the bottom surface. There is a slight groove in the right place, and you may be able to feel a pulse there. Insert an 18 or 20 gauge needle at right angles to the tail, over one of the tail bones (not between) and push in gently until it touches the bone. You should be in the vein, but you may have to search a little from side to side. The tail vein is much smaller than the jugular, but it is still possible to collect 10 mL of blood with a little patience.



### Goats, sheep and horses

Blood can be collected from goats, sheep and horses using the jugular vein, similar to the technique used in cattle and buffalo. In horses, it is done in much the same way, except there is no need to restrain the animal as firmly. Use a halter, or if the animal is nervous, use a twitch as well. Make sure to disinfect the area before collecting blood from horses.

Goats can be bled while standing, if an assistant holds the head up, or else while held down on one side. In both horses and goats, the vein is easy to find.

Sheep can be bled while held on their rumps. The vein can be quite difficult to find, especially if there is a lot of wool around the neck, but the technique is similar.

### Pigs

In pigs there are three places to collect blood. For larger volumes, the cranial vena cava deep in the base of the neck is used to collect blood instead of the jugular. The animal is restrained using a pig snare as described above, and the head is held up as high as possible. The vein can be found on the under side of the neck, just where it joins the front of the chest. Follow the trachea (windpipe) until you feel the bone at the front of the chest (sternum). There is a small indent on either side at this point. Using a 20 gauge 2 inch needle (for adults) push the needle straight into



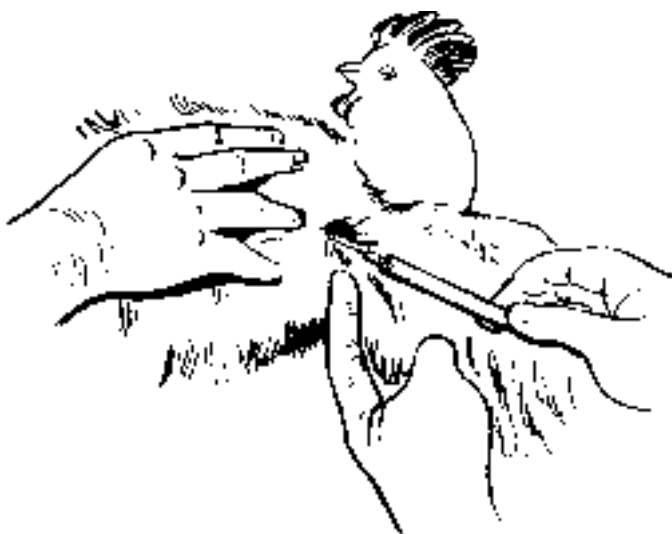
this indent, and angle it slightly towards the middle. It is not possible to see the vein, as it is quite deep, so you will need to search. In small pigs and piglets, it is often easier to do this while holding the pig on its back. An assistant holds the muzzle, and pulls the front legs back along the body.

For smaller amounts of blood, there are two other sites. The ear vein can be used to collect blood, as they are easy to reach, and easy to see. However, the veins are quite small, so a fine needle (19 - 23 gauge) should be used, and not much blood can be collected. You should use a syringe for the ear vein, not an evacuated tube, as the strong suction will collapse the vein.

The tail vein can also be used, if the back end of the animal can be properly restrained. Bleeding is done in a similar way to cattle, although the vein is somewhat harder to find.

## Chickens

Chickens and other poultry are small and easy to restrain, but their veins are small too, as is the amount of blood that can be collected. To collect blood from chickens, use the wing vein, which is located on the underside of the wing and runs beside the first bone in the wing. To identify this, have an assistant hold the bird with its back facing down, and one wing extended so the underside of the wing is exposed. Pluck feathers from the area around the first (elbow) joint. A blue vein is clearly visible, just under the surface. Put pressure on the vein at the base of the wing so it fills with blood and is easier to see. Use a fine needle (20 - 23 gauge) and a syringe (not an evacuated tube) and carefully slip it under the skin and along the vein. Draw back gently and slowly to collect the blood. Alternatively, the needle can be inserted without a syringe, and a tube used to collect the blood as it drips from the hub of the needle. It is very easy to damage the vein and cause bleeding under the skin, which will make it very hard to collect more blood. When finished, put pressure over the site for a few minutes to stop the bleeding. Chickens tend to stop bleeding slowly, and can lose a dangerous amount of blood if pressure isn't put on the site.



## Recording the data

As each specimen is collected, the number of the tube should be recorded on a data recording sheet. There is some other basic information that should also be recorded to assist in the analysis of the specimen. This includes:

- the village from which it was collected;
- the date of collection;
- an identifier for the owner;
- the species of the animal (if more than one species are being sampled in the survey);
- the sex of the animal; and
- the age of the animal.

An example data recording sheet is shown in Appendix D.

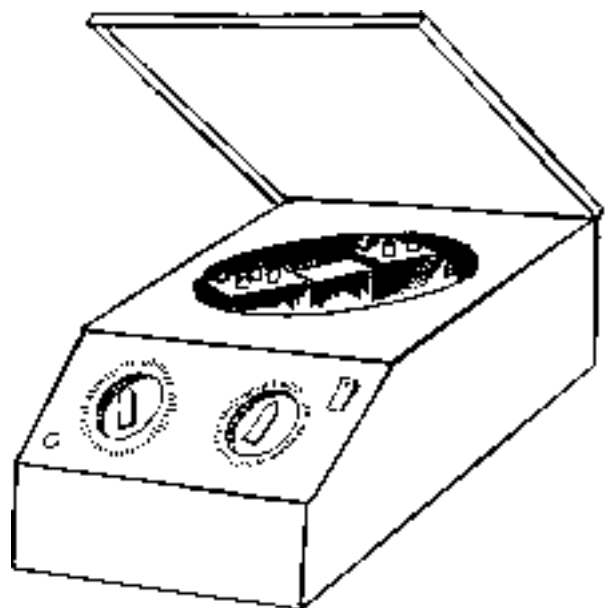
## Transporting and processing blood

When blood is analysed, it is usually the serum or plasma that is examined, although it may sometimes be the blood cells. When the red blood cells are damaged, the red haemoglobin they contain is released. This makes it impossible to use the serum or plasma, as it is stained by the haemoglobin. Blood cells are likely to become damaged if the blood is handled roughly (shaken, or squirted through a needle), is allowed become warm, or is left for too long.

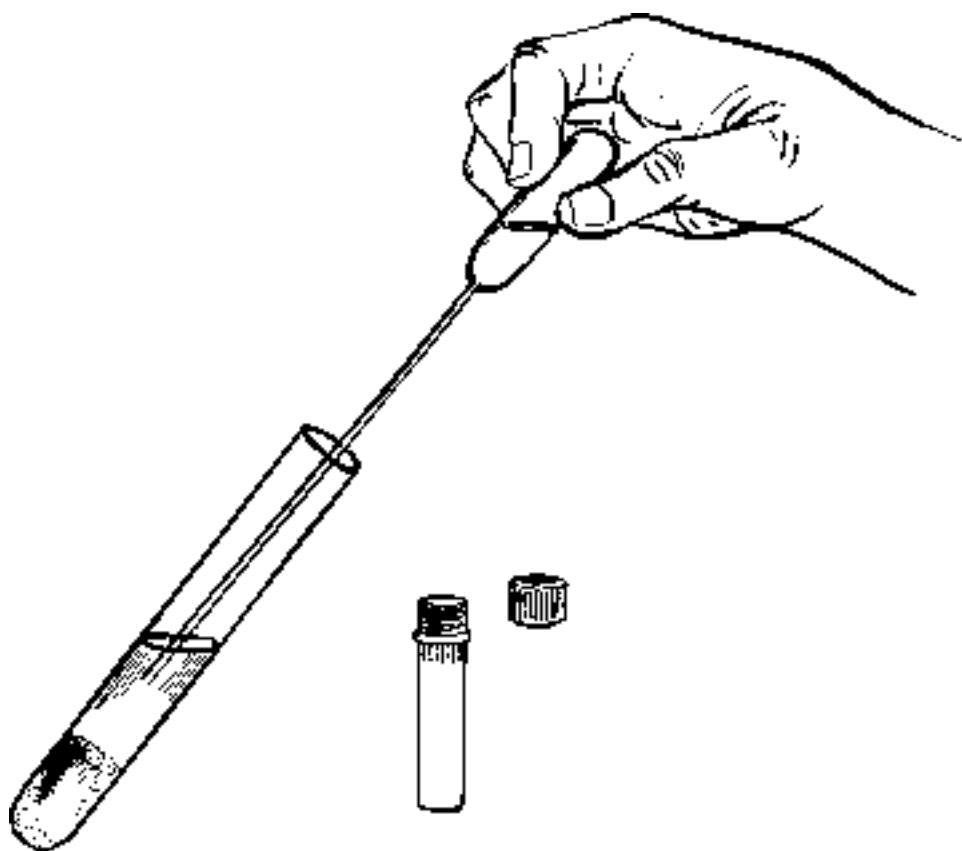
Blood will be in good condition for testing if these three rules are followed; 1) don't treat it roughly, 2) don't let it get hot, and 3) don't let it sit for a long time before testing. When blood is collected in the field, it should be placed straight into a cool box, or car refrigerator. If a car refrigerator is available, this is the best option. If not, a cool box with some sealed ice bricks is good enough to store blood for a single day. Don't use normal ice in the cool box, as it melts, the tubes get wet, and labels run or fall off. Make sure that the tubes are kept upright at all times.

If kept in the refrigerator at about 4°C, blood can be stored for 2 or 3 days without too much damage but it is better to process it as soon as possible.

Blood is made up of *blood cells* (red and white), and the fluid in which the blood cells are carried, the *plasma*. The red blood cells contain the red-coloured haemoglobin which transports oxygen. If not handled properly after collection, the red blood cells can break, releasing the haemoglobin. The red staining of the plasma makes laboratory analysis impossible. The plasma contains many different substances including antibodies, and factors which help the blood clot (clotting factors or coagulants). When



blood is collected, it can be stored in tubes either with or without an anticoagulant (such as lithium heparin, or citrate). If no anticoagulant is present, the blood will clot. The fluid that separates out is called *serum*, and is the plasma without the clotting factors. The best way to make sure that blood is suitable for testing, is to separate it to remove the red blood cells. Once this is done, the serum or plasma can be frozen and stored for a very long time. Use a centrifuge to separate the cells from the serum or plasma. Place the blood tubes evenly in the centrifuge, so it is balanced, and spin them for 10 minutes at about 2000 rpm (or slightly longer for specimens with anticoagulant). When it is finished, all the blood cells should be at the bottom of the tube, and clear serum or plasma at the top. Use a pipette to transfer the serum or plasma from the blood tube to a serum tube. Label the serum tube, and freeze it at  $-20^{\circ}\text{C}$  or colder ready for analysis.



# 5

## Village Interviews

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

General Guidelines

Who should attend

Organising the meeting

Who should lead the interview

Getting good information

Ensuring cooperation

Order of Interview

Introduction and welcome

Building a village sampling frame

Ranking disease priorities

Collecting outbreak history information

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



As discussed in Chapter 4, one of the fastest, most convenient ways of collecting information is to talk to people. Asking people about livestock diseases is much easier and less expensive than collecting blood samples or examining animals. The problem is that the quality of information that is gathered by asking people is often not as good as that collected by direct examination of the animals.

Village or livestock-owner group interviews are a very valuable tool to collect information from livestock owners, and ensure that the quality of the information is as good as possible. Group interviews are faster and more efficient, because all the people you need to ask are gathered together in the one place. The quality of the information is better because you can cross-reference information, comparing the thoughts and ideas of different people. It is easier to get reliable information about events in the past, because livestock owners can help each other to remember things that a single person may easily forget. The group-memory of a village is much greater than the memory of a single individual.

One of the advantages of village interviews is that many different types of information can be collected in a short time. The information collected at an interview depends on the aim of the survey, but some of the key types of information are discussed in this chapter.

These are just some of the reasons for using group interviews. Other benefits will be discussed later. However, in order to achieve these benefits, interviews must be conducted with skill, which takes a certain amount of practice and aptitude.

The type of interview depends on the type of information required. An unstructured interview, which may take the form of a guided discussion of disease problems, is useful to start to understand the main concerns of village livestock owners. A semi-structured interview has a list of topics or areas that need to be discussed, but there is still room for the livestock owners to talk about other issues not on the list. A structured interview has a clear list of specific questions that the livestock owners are asked to answer. Structured interviews are able to quickly collect many different facts and figures. In practice, most village interviews will be a mixture of these different types - specific questions to determine key pieces of information (for instance the date of the last FMD outbreak), and more general unstructured discussions to understand the problems facing the village.

## **General guidelines**

### **Who should attend?**

One of the objectives of a village interview is usually to gather complete information about all the village livestock. Ideally, all owners of the relevant livestock species in the village should attend the interview. For instance, if a survey is designed to collect information about cattle, every household that raises cattle should be represented at the meeting. This is often not possible, but an effort should be made to ensure that as many of the livestock owners as possible are present.

It is usually beyond the control of the survey team to determine which person from the household comes to the meeting. In some societies, such a meeting will be seen as important, and the head of the household, usually a male, will attend. In others, it may be seen as a waste of time, and lower status members of the household asked to go instead. Alternatively, some people may be busy working when the meeting is held, so others have to go instead.

The best people to have at the interview are those that know the animals best. The head of the household may own the animals, but it is perhaps the women or

children who take most responsibility for their day-to-day care, and who therefore have the best knowledge of what diseases they may have had.

The size of the group has an effect on how well the interview can be run. Small groups are easier to manage, but the “group memory” and amount of discussion is also small. Large groups can be very difficult to manage, and the comments of individuals are lost in the general chatter. The ideal size for a group is between 10 and 20 people. When more people attend the interview, one way to make the meeting run more smoothly is to split the group into two or more smaller groups, and run separate interviews. This has the added advantage that the information gathered from one group can be used to check the information gathered from another.

## Organising the meeting

Livestock owners need to know about the interview some time before, so they can arrange to be present. The way in which this is organised depends on the overall organisation of the survey, and the ease of communication with the village.

If communication with the village is easy (either by telephone, or preferably in person), then the head of the village can be contacted a week or two before the planned visit. The head is asked to convene a meeting and invite all the relevant livestock owners. This initial contact is very important, and some brief explanation should be given of the purpose of the survey and the meeting.

Even if the village has been notified a week or two before the meeting, it is often a very good idea to remind the village head one day before. This avoids the problem of forgetting about the meeting.

When access to the village is very difficult and more than one visit is not possible, the first contact with the village may be at the time of the interview. In this case, the survey team should be prepared to notify owners themselves and wait until most are able to come to the meeting.

The best time for the interview depends on the normal activities of the village, and what other activities are planned as part of the survey. Generally, the best time is when it is most convenient for the livestock owners to attend the meeting. An important point, which will be raised again later, is that the owners are assisting the survey by providing the information. Every effort should be made to make their involvement in the survey as convenient as possible and preferably enjoyable. Therefore, meetings should be planned at a time convenient to the livestock owners, not for the survey team’s convenience.

**Example:** A survey is being conducted which involves both village interviews, and the collection of blood from a sample of the village pigs. The survey team could travel to the village early in the morning, and hold the interview before the livestock owners start work. Then the animals could be examined during the rest of the day, and the team could return home in the evening. Another approach would be to travel to the village in the afternoon, hold the interview in the evening, after the owners had finished their work, and collect specimens the next morning, returning home or moving on to the next village during the middle of the day. If livestock owners were happier to have a meeting in the evening, the second option would probably be the best.

Interviews may be held in a community meeting hall, a school, a place of worship, the home of one of the livestock owners or in an open space. Wherever it

is held, it should be quiet, with few distractions, so that the owner's responses can be clearly heard.

## Who should lead the interview?

There is an art to leading an interview, and some people are more suited to it than others. The person that leads the interview should:

- have a very clear understanding of the purpose of the survey, the order of the interview, and the way in which information is to be collected,
- have a good technical knowledge of all the diseases being discussed, and be able to answer questions on these diseases,
- be fluent in the same language and dialect as that spoken in the village. In some situations it is helpful if the leader of the interview even has the same accent,
- be comfortable addressing the group, and able to speak in a clear, loud voice,
- understand and respect the culture and social customs within the village, and make people feel relaxed,
- value and respect the knowledge and skills of the livestock owners and make them know that their knowledge and assistance are appreciated,
- not be intimidating to the livestock owners. They should feel free to express their opinions.
- be able to elicit a response from the quieter or shier members of the group, and encourage participation by all livestock owners.
- be a woman, if interviewing women's groups.

It is often difficult to find one person with all these qualities. When selecting and training survey staff, confident, intelligent, outgoing and sensitive people should be identified, trained and strongly encouraged. A skillful interview leader can have a strong impact on the quality of the information collected, as well as the willingness of livestock owners to participate in future surveys.

## Getting good information

### Listening

One of the major roles of government veterinary officers is extension – providing advice and training on animal health issues to livestock owners. During an interview, some veterinary staff may find it difficult to abandon this role as a provider of information, and have a tendency to interrupt livestock owners to correct misconceptions, provide advice, or, at worst, lecture. The position of the survey team during the interview is that of student, not teacher, and its task is to ask questions and record the answers. For many veterinary officers remembering to listen rather than to speak is one of the hardest things about an interview.

Despite this, a village interview offers an excellent extension opportunity. This is discussed further on page 111 under Ensuring Cooperation.

### Encouraging participation

One of the most important advantages of using village interviews to collect information is the ability to question and collect information from many people at once. If many of the livestock owners at the interview don't participate in the interview (don't offer their opinions or take part in the discussions) then this potential advantage of village interviews is not realised. A successful interview is one in which

all the participants have had the opportunity to express their views and report their experiences fully, and it is up to the survey team, and particularly the interview leader to try to ensure that this happens.

There are many reasons why some people may be reluctant to participate during an interview, but the main one is related to social status and local customs. At most meetings, there will be a range of people from different social levels. Amongst the livestock owners, the head of the village will usually be a highly ranked person, along with other village members with official positions (a veterinary worker, for instance). The other livestock owners will have their own ranking, perhaps related to age, the number of animals kept, or other criteria. In addition to the established social ranking within the village, another layer of ranking is introduced during a meeting to discuss livestock health issues. Those members of the community with more experience or knowledge of health issues will be in a stronger position to participate than those with less knowledge. From this point of view, the survey team themselves will usually be identified as animal health experts, and therefore of very high status. While this may be an advantage, it may also intimidate village members who are uncertain of their own knowledge of animal health issues and may fear exposing ignorance.

The problem with differences in social status is that livestock owners with a higher status are more likely to express their opinions and speak for the rest of the village, while the lower status participants are less likely to speak, and certainly less likely to contradict or correct statements by village leaders. These generalisations do not apply to every society or culture, but the objective of ensuring the participation of all livestock owners still remains. A good understanding of the local culture and a sensitivity to status issues is an advantage, but some other techniques also help.

At the beginning of an interview, there is naturally some initial shyness or reluctance to speak. An activity early in the interview that breaks the ice and gets every livestock owner speaking can help overcome these inhibitions. The form this activity takes depends on the objectives of the interview and what is appropriate according to local customs. It may be a formal part of the interview (such as the collection of information on the number of livestock owned by each participant, discussed on page 115) or could be in the form of a game or activity specifically designed to relax the participants, and get them speaking and thinking about their animals.

**Example:** A competition is a good way to get people involved at the start of the interview. If you are intending to build a village livestock sampling frame, you can use this activity to get people thinking about how many animals they and their neighbours have. Divide the livestock owners into several teams (4 or 5 if there are enough people). Ask each team to try to calculate the total number of livestock (of the relevant species) in the village. Tell them that the team that gets closest to the true value will win a prize. Give them 5 minutes to think about how many animals there are, then ask each team to report its answer. Write the teams and their guesses on a board or large sheet of paper. You can then go ahead with building the sampling frame (page 115) to work out the real total number of animals. The prize can be several packets of worm drench or some other simple item which assists with the care of the animals.

This type of exercise starts people thinking about the number of livestock that each owner keeps, including those who may not be present at the meeting. It also makes the next activity (asking each owner how many animals they have) much easier to understand – it is being done to find out who won the competition. It also underlines the fact that it is really just the total number of animals that is required, and information on the number kept by each owner is not going to be used for any purpose other than to calculate that total.

Another way to make it easier for all livestock owners to participate is to try to minimise the perceived difference in status between the survey team and the livestock owners. This can be done by physically reducing the distance between the two. In meetings where participants sit on the floor or ground, the survey team should also sit on the floor or ground. In any case, where the size of the group permits, sitting in a circle is better than the traditional “speaker - audience” seating arrangement. The distance can be reduced verbally as well, by letting the livestock owners know that you understand their problems and practical difficulties.

Directing specific questions to specific members of the group can also help to encourage participation. Often a few participants will emerge who tend to do most of the talking. If a general question is posed, and one of these more vocal people answers it, it can then be asked again, directed at specific people.

**Example:** One of the objectives of an interview is to determine which is the most important disease amongst pigs in the village. This is done in two parts. First, all the diseases that occur in pigs in the village are listed. Then, the group identifies which is the most important. During the interview, the question is put to the whole group, and a village leader answers that deaths amongst young pigs is the most important problem. The leader of the interview then turns to one of the less vocal livestock owners and asks “Do you have a problem with deaths amongst young pigs?”, “Do you have any other problems with your pigs?”, and “Which of these problems is most important?”. This can be repeated for several of the other owners, to confirm the initial response, or to get a better picture of what diseases are occurring and how the livestock owners rank them in importance.

One group in many societies that often has a lower social status is women. This is perhaps even more noticeable when it comes to raising livestock (particularly large animals) which may be seen as a male activity. Despite perceptions about their lower status, it is very important to encourage women to attend the interview, and to stimulate them to participate. There are several reasons for this. While women may not be thought of as the owners of the livestock, they are often responsible for most of the care of the animals, and may be the people that spend the most time with the animals. For lower-status animals (such as poultry), they may have sole responsibility for them. As such, women have a great deal of knowledge of the problems that affect the animals. Another reason for encouraging the participation of women is because they represent a separate social network within the village, and are interested in, and have access to, different types of information to the men. For example, when building a village sampling frame (page 115), women are often able to provide better details on the number of animals kept by livestock owners who are not present at the meeting.

In many societies, women will sit in a separate group to the men during a meeting. If there are only a few women present, it may be difficult for them to make

a contribution. However, if enough women are present, questions may stimulate some discussion amongst the women's group. A more confident spokesperson may emerge from the women's group to report on these discussions, but if not, targeted questions to particular women should be able to discover their perspective on the issues being discussed.



### **Language**

The language used during an interview plays an important role in the participation of the livestock owners and the quality of information collected. Clearly, the leader of the interview, and the person recording details should be fluent in the local language and dialect. Speaking the same language and using the same expressions helps to reduce the distance between the survey team and the livestock owners, and encourages better participation.

The language used should also be kept clear and simple, and avoid using technical terms (something that trained veterinary staff sometimes find difficult). Nevertheless, there is a fine line between using simple language and appearing to be condescending.

### **Disease Names**

The choice of words is particularly important when discussing particular diseases. Veterinary staff may be interested in a particular group of diseases, which they usually refer to by their technical name (often in English). The survey team will often think about diseases as particular entities, each with a particular separate cause. On the other hand, livestock owners may not be aware of the specific causes of particular diseases, and think more in terms of disease syndromes. When an animal behaves in a particular way, and shows particular signs, then it is thought of as a particular disease. This disease syndrome may have a unique local name, or the technical name of a particular disease may be used.

**Example:** In a particular village, the livestock owners identify a particular disease as being very important. They describe the disease as affecting cattle and buffalo, causing a high fever, loss of appetite, slight bloat, and death in most cases after 2 or 3 days. They have a local name for this disease, but when asked by the survey team, they call it haemorrhagic septicaemia. The local veterinary officer has visited the village to explain about haemorrhagic septicaemia vaccination, and said that this was the technical name of the disease. There has never been any post-mortem or laboratory confirmation of the cause of these deaths, and in fact, there are several different diseases that occur in the village, all appearing similar, that are grouped together into the same syndrome. haemorrhagic septicaemia is only one of these diseases.

The tendency of livestock owners to talk about disease syndromes based on patterns of clinical signs, rather than specific diseases should be kept in mind during the interview. Mistakes can be avoided by paying attention to a few points.

Don't use the technical name for a disease when asking about that disease. If you know the local name, and understand what disease or diseases it truly represents, then use that name. Better still, if interested in a particular disease, describe the signs of that disease, or show pictures of animals showing typical signs, and ask the owners to tell you what it is called. Explore whether this name is used for a single disease, or if other diseases may be grouped together. If livestock owners use a technical name, question them to check that you are both talking about the same disease.

In some situations, the clinical signs and behaviour of the disease in the population are distinctive enough to be sure that the name given by owners does refer to a single disease. In others, it is not so simple.

When collecting and analysing disease information from livestock owners you need to remain aware of the type of information that is being collected and its quality. When a disease syndrome is described, which may include several other diseases, it is not possible to distinguish between the different possible causes in the analysis. If a survey was conducted to assess the incidence rate of haemorrhagic septicaemia based on interview data, but the local name for the disease included several other diseases causing sudden death, it would not be possible to report that the incidence rate of village outbreaks of HS was, for example, 12 per 100 villages per year, but rather that the incidence rate of outbreaks of disease causing fever and rapid death was 12 per 100 villages per year.

### **Persistent questioning**

One rule of collecting information through interviews is that you should never be satisfied with the first answer. When a question is asked, there is the danger that the answer could be wrong, either for the reasons listed previously, or else because the experience of the person answering doesn't represent the experience of the whole village. It is a good idea, therefore, to check and recheck every answer that is received. This is done by asking the same question in several different ways, to several different people. Each time, the question focuses on some different aspect of the problem, and each answer is compared. If there is some inconsistency, then a discussion is started to try to resolve it and come up with a consensus.

**Example:** A list of all livestock owners is being built to act as a sampling frame. All those present have reported the number of animals owned. The

group is asked “Are there any livestock owners who aren’t here at the meeting?”. The group responds with three more names. “Is that all? Are there any more?” One person responds that there aren’t any more. “Does anybody have a neighbour with animals who is not here?” One person realises that they do, and their details are recorded. “Are there any people who live outside the village and keep animals?”, “How about on the road leading to the north?”, “On the road leading west?”

This type of persistent questioning, and prompting to help the owners remember information, can be continued until the survey team are convinced that they have gathered the best information possible.

## Ensuring cooperation

The cooperation and goodwill of livestock owners is essential for a successful survey. For many disease surveys, the only people that hold the information required are the livestock owners, and the only way to collect specimens is from their animals. If owners are unwilling to cooperate, then the most valuable source of information on the diseases of livestock has been lost. Without this information, any attempts to control these diseases may be much more difficult. The key role of the owners means that every effort should be made to ensure that they are happy to participate in the interview, and (if necessary) are happy to allow the survey team to collect specimens from their animals.

If no livestock disease survey has been conducted in the village before, most owners are likely to be reasonably happy to cooperate, perhaps more out of curiosity than anything else. However, the aim of the survey team should be to make sure that the owners are happy to cooperate the *next* time a survey is conducted as well. Although there may be no plans to survey this village again in the near future, any ongoing disease control program will need to be monitored with regular surveys. If, at each survey, the livestock owners become unwilling to participate again, then as time goes by, it will be more and more difficult to find cooperative villages.

The problem with ensuring future cooperation is that livestock disease surveys are mostly designed to take from the village, not to give. During a survey, information and blood samples are collected, and the survey team leaves to move on to the next village. The benefits to the survey team are very great – they have information and specimens necessary to understand the disease situation, and help manage disease control programs that will benefit the entire country. The benefits to the village and the livestock owners that gave the information and specimens are not so clear. When the survey team leaves, all that has happened is they wasted a few hours at an interview, and had to help hold their animals while blood was collected, stopping them from working, and upsetting their animals. It is clear that some livestock owners can see no direct benefit to themselves and could be reluctant to participate in future. The challenge to the survey team is to provide some direct benefit to the owners so they will be happy to assist next time. The best way to achieve this will depend on the situation and culture, but a few suggestions follow.

### Explaining the objective of the survey

One of the easiest ways to make owners more cooperative is based simply on good communication. It is probably one area which is most often forgotten. At the start of the interview, the leader should carefully explain the purpose of the survey, the village’s role in the survey, and the benefits the village will get from the survey. This should include the following points:

- The survey is conducted by the government (or other organisation) to collect information which will help solve disease problems throughout the country. (The objective of the survey can also be explained. For instance to collect information to allow the government to decide which diseases are most important, so they can allocate more funds to trying to solve these disease problems.) The results of this survey will therefore benefit all the livestock owners in the country (or province etc.), not just this village, but they may not be noticeable for a while.
- The survey is not working to directly assist *this* village or the other survey villages, but *all* villages.
- The village has been chosen at random to be representative of all the villages in the area.
- The information collected will only be used to try to solve animal health problems, and won't be given to anybody other than the veterinary authorities. (This is to allay the fear of some owners that information may be used for taxation or other purposes).
- Explain how long the interview is expected to last, and what is expected of livestock owners after the interview (if specimens are to be collected).

If the livestock owners' expectations of the survey are realistic (i.e. that it is not aiming to directly benefit their village), then they are less likely to be disappointed. If they are made aware of the importance and potential benefits of the survey at the wider level, they may be happier to assist for the general good.

### **Attitude**

The attitude of the survey team towards the livestock owners will influence the way they feel about helping with the survey. The survey team should realise that the owners are experts on their own animals and their health, and that the information they hold is very important. If the owners realise that their opinions and experience are respected, and that the team is grateful for their help, they are likely to feel better about participating, and to be proud of their contribution.

### **Payment**

Despite good explanations and a demonstration of respect towards the livestock owners, it is still plain that there is no direct material benefit to them from the survey. In some circumstances, it may be necessary to provide some sort of payment to livestock owners, so that there is some benefit for the participation in the survey.

This payment may take the form of money, perhaps as an inducement to come to the interview, or as a per-specimen payment for blood collection. While it is sometimes necessary, cash payments should be avoided where possible, for two reasons. Firstly, it makes the survey more expensive, and the veterinary authorities of developing countries can rarely afford such extra expenses. Secondly, it builds expectations amongst the livestock owners. If, during a survey, payment was given to participating farmers, any future survey of that village which is unable to provide payment is very unlikely to get cooperation.

While cash payments are generally not a good idea, payments in other forms may be much more acceptable. For instance, the distribution of free worm treatments to cattle owners as a way to thank them for their assistance will usually build good will, without the problems of cash (such as creating expectations and decreasing cooperation in future surveys if no payment is made then). Other types of payment

in kind, that are able to directly benefit the health of village animals can also be used (e.g. free vaccination). Consider, however, that if private veterinary workers supply medicine to the village, (unsustainably) handing out free worm treatment or vaccination may undermine the (sustainable) private system, and should therefore be avoided.

### **Information**

The veterinary services of most developing countries may not have a lot of money to pay livestock owners, but they do have information on diseases, and this is one way to provide real benefits to the livestock owners, in payment for their participation. A village interview is a good opportunity for the survey team to collect a lot of information from the livestock owners, but it is also a very good opportunity for the livestock owners to collect information from the survey team. For many villages, it may be quite uncommon for skilled veterinary staff to visit. After a village interview where disease problems have been discussed, livestock owners are likely to be thinking about all the problems they have had with their animals. The survey team is able to provide all the advice and information that the owners require.

Providing information and answering livestock owners' questions is therefore an important way to give some benefit to the owners. This is best done as one of the last parts of the village interview, when all relevant disease issues have been discussed, and all the necessary information collected. Livestock owners are then invited to ask any question relating to the health of their animals, for discussion or advice. If necessary some topics could be suggested that have arisen out of the interview.

This part of the interview may not gather any more information for the survey, but should nevertheless be seen as an important component. As much time as required should be spent to address all the questions raised. Veterinary staff used to making lecture-style extension presentations should guard against this, and be careful to listen to the question, and address it specifically rather than embarking on a long, dry, prepared presentation on a topic.

### **Fun**

If participants in the survey enjoy the experience, for whatever reason, they are likely to feel better about helping with the survey. One way to do this is to provide some form of entertainment or recreation as part of the survey visit. There are many ways that this could be done, and it may achieve multiple purposes.

At the simplest level, all participants could be invited to a meal at the end of the interview, provided by the survey team. Enjoyable for the livestock owners, this is also an opportunity to discuss issues in a less formal setting, and get a better understanding of the disease problems in the village. If the survey team stay overnight, this idea could be extended to a meal and a party. Some other form of entertainment could be provided, that also serves an extension purpose. For instance, a show could be presented, conveying some important animal health themes, but in an entertaining form. Alternatively, a video or film with the same objectives could be prepared and shown during or after the interview. For this to be successful, the video must be entertaining, and not a dull lecture on disease control.

**Example:** The veterinary services of one country have commissioned a television studio and cast of a popular television show to produce a special episode dealing with animal health issues. The characters are well known throughout the country, and the plot, as with every episode, is very

entertaining. This episode is shown by survey staff at the end of the village interview, both to entertain, and to underline important messages.

## Order of interview

The structure of a village interview depends on the types of information being collected. It is important to deal with some issues before others. For instance, questions which aim to discover what diseases occur in the village, or what the owners' understanding of or names for diseases are, should always come before questions on a specific disease.

An example of the structure of an interview as part of a village survey is shown below. This interview could be used for a village survey which aimed to collect several different types of information simultaneously – blood specimens from cattle and buffalo for estimating the prevalence of antibodies to foot and mouth disease; information from livestock owners that would assist in setting priorities for disease control; and information to estimate the incidence rate of foot and mouth disease outbreaks in the past. The various activities and outputs of the interview relating to sampling and the different survey types (prevalence and incidence rate) described in this book are discussed in detail below.

## Activities and outputs

### Activities

- Activity 1: Village interview
  - Activity 1.1: Introduction and welcome
  - Activity 1.2: Build a sampling frame
  - Activity 1.3: Select animals for blood collection
  - Activity 1.4: List and rank major disease problems
  - Activity 1.5: Collecting outbreak history information (e.g. FMD)
  - Activity 1.6: Open discussion and answer questions
- Activity 2: Blood specimens
  - Activity 2.1: Collect blood specimens
  - Activity 2.2: Record data
  - Activity 2.3: Process and store specimens

### Outputs

- Output 1: Sampling frame
- Output 2: List of animals for selection
- Output 3: Completed interview form
  - Output 3.1: Major disease problems ranked by importance
  - Output 3.2: Date of the last FMD outbreak
- Output 4: Blood specimens and record sheet

### Introduction and welcome (Activity 1.1)

The beginning of the meeting is a good time to explain the objective of the survey, as mentioned on page 111. At the same time, the survey staff should be introduced, and the activities within the village explained. In any survey involving livestock owner participation, various problems are sure to arise, many due to

misunderstandings or failure of communication. The best way to deal with these problems is to address them before they develop. Experience during trial surveys or previous field work will indicate where the problems are likely to be. Here are a few examples of potential problems, and ways to address them during the introduction to the meeting.

- Why was this village chosen? Why do you want to collect blood from my animals? Some livestock owners may feel suspicious or victimised that they have been singled out to participate in the survey. There may be a feeling that the survey team is implying that their animals are more diseased than those of their neighbours. To avoid this impression, it should be explained how the village or individual animals were selected. For villagers with no experience of sampling or probability, the concept of random selection may take too long to explain. One approach is to explain that the computer made the choice of which village and which animal to select. As a machine, rather than a person made the decision, there can be no sinister intent, as the machine clearly knows nothing about them. Alternatively, a simple explanation using the analogy of dice, cards or the lottery could be used.
- Are you going to use information on the number of animals for taxation purposes? People are often cautious about answering questions regarding their wealth, especially when the question comes from somebody seen to represent the government. In many societies, one measure of wealth is the number of animals owned. Explaining how the information will be used should remove any fears. For instance, when building a sampling frame, it could be explained that the survey team is only interested in knowing how many animals there are. The information on the owner is just there so the team can identify the animals. After the animals are sampled, the information on the name of the owner is thrown out.
- Will collecting a blood specimen hurt the animals? In many villages, the only experience livestock owners have had with needles is for the injection of vaccines or other medicines. Where vaccination campaigns have been running for some time, effective extension may have given the message that you shouldn't vaccinate pregnant animals or young animals, with the risk of causing abortion or some sort of harm. The difference between collecting a blood sample and injecting vaccine may not be clear, and farmers may therefore be concerned about letting the survey team sample young or pregnant animals. To address this problem, it is necessary to explain the difference between vaccinating and collecting a blood specimen, and to assure owners that it will do their animals no harm at all.

## **Building a village sampling frame (Activity 1.2)**

A key reason for holding a village interview is to quickly and easily collect information for a sampling frame, so that a random sample of animals can be selected. The process is straight-forward, but care is needed to ensure that the sampling frame is as complete and accurate as possible.

The best way to make sure of the completeness of the information is to have as many livestock owners present at the interview as possible. This is mainly dependent on the way the village was notified about the meeting, how its purpose was explained, and whether a reminder was given. If it was not possible to tell the

livestock owners about the meeting well before, it may take some time and effort to collect most of them together.

Building a sampling frame should be one of the first activities in the interview, immediately after the introduction. One advantage is that it helps break the ice, and encourages people to speak. Each person at the interview is asked a direct question and is required to answer out loud.

The process of building a sampling frame is:

- Step 1:** Explain the activity.
- Step 2:** Ask each person, one by one, for their name, and the number of all relevant species that they (or their household) keep. For instance, if a survey aims to collect a random sample of goats, only the number of goats kept by each household is needed.
- Step 3:** Record the responses on the numbered data recording sheet, as shown in the example below. A full copy of the data recording sheet is included in Appendix D.

N°	Name	Goats	Cum. Total	Selected
1	Lung Noi	5		
2	Tu Nyai	5		
3	Silipak	4		
4	Khamphone	10		

- Step 4:** When each person present has been asked, ask if there are any other livestock owners who are not at the meeting.
- Step 5:** For each of the livestock owners who is not present, have the group estimate the number of animals kept.
- Step 6:** Keep asking until you are sure that nobody has been missed (see Persistent Questioning on page 110).
- Step 7:** One or two members of the survey team can then use the sampling frame to select animals to sample while the rest of the meeting continues. The process of selection from a village sampling frame (Activity 1.3) is explained in detail in Chapter 3 on page 55.

## Ranking disease priorities (Activity 1.4)

This section provides an example of just one of the many different types of information that can be collected during a village interview. When deciding on priority diseases for control programs and government expenditure, many considerations must be taken into account. Amongst these are the overall costs of the disease, which are related to its prevalence or incidence rate and the loss associated with a case of disease. Other criteria for determining the priority of a disease include its importance to international trade, and to public health. One consideration should always be recorded is the importance of the disease to livestock owners. This information is easy to collect quickly during a village interview.

Seasonal patterns of disease

The only way to effectively control a disease, and to control it as efficiently as possible, is to understand the disease first. This involves knowing not only which diseases are occurring and where, but also when they are occurring. An

understanding of the seasonal pattern of disease occurrence can lead to the development of better targeted control measures.

**Example:** A locally available haemorrhagic septicaemia vaccine is able to provide protection for only 6 months. The cost of vaccinating all animals twice yearly is too high, and the veterinary authorities are faced with a choice of either vaccinating fewer animals, or vaccinating them only once a year. An examination of the patterns of disease indicates that 90% of haemorrhagic septicaemia outbreaks occur in a five-month period during the wet season. If animals are vaccinated once a year, at the beginning of the wet season, 90% of outbreaks could be avoided, at only half the cost of vaccinating twice yearly.

When collecting information to rank diseases from the livestock owners’ point of view, it is also easy to collect information on the patterns of those diseases in the village. To collect this information:

Disease ranking

- Step 1:** For each of the animal species of interest (or for all species), ask the group to list all the diseases that occur in the village. Make sure that the names used in the village are recorded, as well as a description of the signs of that disease (see Disease Names on page 109).
- Step 2:** As each disease is described, ask in which months the disease *usually* occurs. Use an appropriate time scale for the village – some will be comfortable to report information in calendar months, while others may prefer to use an agricultural calendar, reference to ceremonies, or seasons to describe different times of the year. Use whichever is most commonly used in the village, but be sure that you are able to relate the times back to calendar months.
- Step 3:** Record the information on a data recording sheet, such as that shown below. Use a line to indicate that diseases occur over a period of months. A full copy is included in Appendix D. A large copy should be made and displayed so all the people can see it, for instance on a large sheet of paper against a wall.

Disease Name	Description	Species	Rank	J	F	M	A	M	J	J	A	S	O	M	D
FMD		Cattle													
Diarrhoea		Pigs													
Sudden Death		Chickens													

**Step 4:** Ask the owners to rank the diseases in order of importance.

Ranking techniques

There are several ways of ranking, and many criteria for what makes a disease important. In the simplest situation, just ask the village to discuss the diseases and decide which are the first, second and third most important diseases overall, based on any criteria they care to use.

A much more complex and time consuming approach is to do a full ranking exercise. This involves discussions to establish which criteria should be used to judge the importance of a disease (for example, number of deaths, cost of treatment, loss of work, loss of weight). Then for each of these criteria, each disease is given a score to measure its importance. For illiterate livestock owners, scoring could be done with pebbles or beans. For each different criterion, the group is given 20 beans, and asked to place them on the different diseases according to the rank. This is done by the group as a whole.

Simpler techniques using some of these methods can also be used, depending on the level of detail required in the information. A number of points should be kept in mind for all questions to determine disease priorities:

- Don't mention specific diseases before the ranking exercise. Even if the survey aims to collect information about one disease in particular, there should be no mention of this disease until after the disease ranking is finished. This includes the first contact with the village to notify them of the meeting. The reason is that livestock owners are likely to think that the disease being studied *must* be important if the survey team has come to study it, even if they wouldn't normally consider it.
- For the same reason, the leader of the interview shouldn't prompt the livestock owners with diseases that may be important. No specific diseases should be mentioned other than by the livestock owners. If they need help getting started, rephrase the question, or else give an example for other species that aren't relevant to the survey (e.g. dogs).
- Don't interpret the responses. When recording data, use local names for disease syndromes, rather than technical names for specific diseases. If necessary, an assessment may be made later as to which diseases a syndrom may represent.
- Don't comment. When owners are listing diseases, the role of the leader of the interview is to listen, and to help them remember different diseases. There is an opportunity at the end of the interview to correct any misconceptions or respond to questions.

## Collecting outbreak history information (Activity 1.5)

Measuring disease incidence rate is one way of assessing the level of disease in the population. Reduction in incidence rate in response to a disease control program is an indicator of the success of the program. Traditional measures of incidence rate require complete reporting over a reasonably long period, which is very hard to achieve in a developing country. Chapter 8 describes a technique to estimate the incidence rate of village disease outbreaks, based on information collected during village interviews. The key piece of information required is the time of the most recent outbreak of the disease in that village.

Strengths and weaknesses

This is an example of asking questions about events in the past. People often find it difficult to remember details of things that happened a long time ago. This approach to estimating disease incidence rate has the advantage that it is fast, simple and inexpensive, but it also has the problem of depending on the memory of livestock owners, and may therefore be inaccurate. To overcome this problem, there are several techniques that can be used to ensure that the answers provided are as accurate as possible.

Firstly, the technique should be used to collect information only about particular types of disease. These are diseases which occur as recurrent outbreaks within villages, which have easily recognised signs, and which have a large impact. The larger the impact at the time of the outbreak, the easier it is to remember the outbreak years later. For instance, a disease that causes animals to die is easier to remember than a disease that just makes them sick. On the other hand, a disease that makes the owner work much harder for a long time (e.g. nursing the animals, cutting feed, administering treatments) may be a lot easier to remember than something that came and went quickly (such as the sudden death of an apparently healthy animal). Disease outbreaks which have a significant financial impact, causing a loss of money, are also more easily remembered.

Another way to make remembering disease outbreaks easier is to ask groups rather than individuals to try to remember them, hence the use of village interviews. During a village outbreak, some livestock owners may be severely affected, while others suffer no losses at all. If you ask the ones who didn't have any sick animals, they may have forgotten the outbreak. If you ask the ones with animals that died, they are much more likely to remember it. Using a group interview taps into the collective memory of a group of people, who can compare their memories with their neighbours to try to determine when the real date was.

The process of collecting this information is as follows:

- Step 1:** Make sure livestock owners are familiar with the disease you are interested in. If it has already been mentioned and described during the interview, this is not necessary. Otherwise, without mentioning its technical name, describe the signs of the disease, and ask the owners if they recognise the disease or know about it. Pictures of clinical cases may be very useful.
- Step 2:** Ask if there has ever been an outbreak of the disease in this village.
- Step 3:** If there has been an outbreak, ask the owners when the most recent outbreak first started.
- Step 4:** *If it is difficult to remember the year, help them by building a village history (see below).*
- Step 5:** Once the year has been established, try to determine what month the first animal got sick. Use calendar months, an agricultural calendar, or seasons, as appropriate, but try to collect the date as accurately as possible. *If they are having trouble, build a village calendar to help (see below).*
- Step 6:** If the owners say that there has never been an outbreak, you need to find out what that means. There may well have been an outbreak many years ago, that everybody has forgotten. You need to try to find out the earliest date since which the villagers are sure there hasn't been an outbreak. For instance, if they say they can't remember an outbreak, but say that if there had been one more than 5 years ago, they wouldn't remember it, then the time to record is the date 5 years ago.
- Step 7:** Record the information on a data recording sheet, such as that shown in Appendix D.

**Village history**

Unless the outbreak occurred recently, remembering the year of the outbreak is the hardest thing. It is easy to make a mistake and say that the outbreak happened one or two years earlier or later than it really did.

Landmarks

A village history can be used to help avoid this problem. When people remember things, they use *time landmarks* to identify when they occurred. These landmarks are events that they can remember clearly, and know the date of. They can then work out whether the event of interest happened before or after the landmark. By using a series of landmarks, and deciding if the event happened before or after each, it is possible to determine the time of the event reasonably accurately.

People use this method naturally when trying to remember things. During a village interview, it is possible to help people remember by encouraging the process and providing landmarks. The landmarks are significant events in the history of the village, that happened at known times, and are easy for everybody to remember.

Building a village history

To build a village history, ask the livestock owners about recent important events that they can all remember, and that happened at a known time. These events may have taken place in the village, or be external to the village. For example, the building of a new temple, or election of a new village head might be suitable landmarks. Natural disasters like floods, fires or storms could be used, as well as national or international events, like elections, significant sporting events, or wars. All these things can be reliably dated, and used as reference points to help identify the year of the outbreak.

**Village calendar**

Once the year has been established, the aim is to determine the month of the start of the outbreak. This is usually somewhat easier, as the agricultural activities at the time of the outbreak can usually be remembered. However, when there is uncertainty, a similar technique can be used to determine the correct month.

A village calendar is made up of landmark events which occur every year at about the same time. These events are recorded next to a calendar showing the months, and livestock owners asked to think if it happened before or after this or that event. Examples of suitable events for a village calendar include events from the agricultural calendar (ploughing, harvesting etc.) and religious festivals. Using a combination of these, it is usually possible to determine which month the outbreak started in.

**Villages with no outbreaks**

When livestock owners report that there has never been an outbreak in the village, it provides us with important information. However, there is the possibility that the owners are wrong, and that there has been an outbreak some time ago, that they have forgotten.

When analysing the results, if we simply say that there has never been an outbreak, we might be wrong, and give incorrect results. On the other hand, if we ignore these villages, we are ignoring important information, and the results will be wrong again.

Censoring

The solution is to use just as much information as we can get. We can be confident that a village reporting no outbreaks hasn't had an outbreak for a certain period of time. Before that time, we don't know what happened, but we do know there has been no outbreak since that time. In the analysis of this information, the time is called the *censoring* time, because past that time, the information is censored or cut off.

In order to analyse the data correctly, we therefore need to know the time since which the owners are sure there has been no outbreak. This idea is sometimes hard to explain, but examples or prompts may help. One technique is to suggest some times. For instance "Are you sure there has been no outbreak in the past 3 years?". If they answer yes, then "Are you sure there has been no outbreak in the past 30 years?". They will usually answer no, because they can't remember what happened 30 years ago. Continue this until they are able to suggest the longest time since which they are sure.

An alternative approach is to identify the oldest livestock owner present. Ask this person how long they have owned and been working with animals. If it has been all their life, ask them if they think they would remember an outbreak if there had been one when they were young, say about 20 years old. If they say they would, then you can use their age minus 20 years as the longest time since which they are sure.

# 6

## Computerised Data Management and Analysis

### Contents

Chapter 1: Introduction

#### Part I: Background to Disease Surveys

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

- Principles of data management and analysis
  - Hardware
  - Software
  - Data
- Data processing procedures
  - Initial check for completeness and accuracy
  - Data coding
  - Creation of a table
  - Data entry
  - Checking for errors after data entry
  - Recoding
  - Linking data
  - Converting data between different formats

#### Part II: Survey Design and Analysis

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### Part III: Notes for Trainers

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



## Use of computers

For small livestock disease surveys using only simple analysis, all the calculations can be done by hand or a simple calculator. However, for large surveys, or surveys with more complex analysis, the amount of data that must be managed, and the types of calculations make it much too difficult to be done by hand. Computers make managing and analysing large amounts of data much faster and easier, and allow some types of complex analysis that would otherwise be impossible.

Using computers has other advantages. Once data have been entered into a computer, many different types of analysis can be performed and reports generated without the need to re-enter the data. When performing analysis using a computer, there is no need to remember the complex statistical formulae, as they are coded into the computer program. Computers make it much easier to avoid and correct mistakes as well.

Computers are therefore important tools for the livestock disease surveys described in this book. The accompanying disk contains all the specialised software needed to carry out the particular types of analysis required for these surveys. Only a basic familiarity with computers is required to use these programs. However, for more general data management and analysis, other software is required. A wide range of suitable programs exists, and it is best to use whichever you are already familiar with. This chapter discusses some general aspects of data management analysis, and makes particular reference to the Epi Info program. This program has several very clear advantages: it has the capacity to perform all the data storage and management tasks required, it is able to carry out a wide range of standard statistical procedures and specialised epidemiological analyses, and it is available free of charge.

## Principles of data management and analysis

This section provides a very brief overview of the use of computers for data management and analysis. If you have experience in the use of computers and computer databases, you may choose to skip to the next section.

## Data and information

The aim of data analysis is to convert a large amount of *data* collected during the survey into a small amount of meaningful *information*. In effect, analysis tells us what the survey data actually mean. To achieve this, the data are converted into a few easy-to-understand measures.

A computer is a tool to help us convert data into information, and has the main advantage that it can process a very large amount of information very quickly. When we use a computer, we are working with three separate components: the *hardware*, *software* and *data*.

### Hardware

Hardware describes the computer itself. The computer is made up of several components with different roles.

First, there must be a way of putting data into the computer, or *data input*. The keyboard is the main way this is done, through typing the data in. There also needs to be a way of storing the data once they have been entered. Storage, or computer memory, comes in two forms. Disks (such as the hard disk inside the computer, floppy disks, or CDs) can store data for a long time, and require no power. In contrast, the memory inside the computer (*RAM* or Random Access Memory) is able

to store information only while the computer is turned on, and this memory is used for temporary storage and doing calculations.

The most important part of the computer is the part that does the calculations. The *central processing unit* (CPU) is the 'brain' of the computer, and is a small silicon chip capable of performing a very large number of calculations every second.

Once the data have been entered, stored and analysed, we then need to know what the results are. The parts of the computer for sending information out are called the *output devices*, and include the screen or monitor, and the printer.

## Software

*Software* is the general name for computer programs. A program is a set of instructions which the computer is able to read, that tells it how to process data. For instance, the computer hardware doesn't know anything about statistics. To calculate statistics, a statistical program must be loaded first. The program has a set of instructions that tells the computer how to calculate the statistics. To calculate the average the program tells the computer to add up a list of numbers, and then divide the sum by the number of items in the list.

Types of software

There is an enormous range of software available, designed for an enormous number of different purposes. The commonly used software can be divided into several main purposes:

- Word processors. These are designed for working with words: writing letters, reports and other documents. They allow the computer to be used like a very sophisticated typewriter.
- Spreadsheets. These programs are for carrying out mathematical calculations and are widely used for business purposes.
- Databases. A database is a program for managing large amounts of similar data. Databases are the main program to use for storing and managing the results of the livestock disease surveys described in this book.
- Statistical programs. These programs use information that has been stored in a database, and perform a range of statistical calculations on those data.

Some programs are able to perform several of these functions. For instance, Epi Info can be used as a word processor, database and statistical program. Others have very specific functions, like the programs provided with this book. Each program performs just one specialised calculation or statistical analysis.

---

**Epi Info**

---

This chapter contains advice which is applicable to any computer program. However, examples and specific instructions are provided for carrying out procedures using Epi Info, as this should be available to all readers, and has the ability to perform all the tasks required. Where instructions are given on using Epi Info, the paragraph is marked with the symbol shown on the left. Only very brief explanations of the Epi Info commands are given. For more details, use the on-line Epi Info manual.

## Data

The third thing that is necessary when using a computer are the data. This is what the computer works with. With livestock disease surveys, the data are all the facts and figures that are collected during the field work, or else produced by the laboratory when analysing specimens.

### Data types

There are many different types of data, but most of these can be represented in the computer in just a few different ways.

- |   |   |
|---|---|
| Text  | • Text. In computer language, words are called text or strings. A written description of something represents data stored as text. For example, village names, the names of livestock owners, the names of diseases that are important in a particular village are all text.  |
| Yes / No                                    | • Yes/No. Answers to questions can often only be yes or no. Many different types of data from livestock surveys can be thought of as having yes/no answers. For instance “Has there ever been an outbreak of swine fever in this village?”. Other types of information that only has two states can also be thought of as yes/no data. When analysing blood for antibodies, the results can be reported as Yes (antibodies present) or No (no antibodies detected).   |
| Numbers                                     | • Numbers. Numbers are used to describe many different types of data. Data stored as numbers can be divided into two groups:  |
| Continuous data<br>Quantitative information | • Continuous data uses numbers to measure the value or quantity of something ( <i>quantitative</i> information). Continuous data may take any value within a range. Examples of continuous data include age, weight, temperature, or livestock population. Continuous data may be represented as <i>integers</i> (whole numbers, such as the number of pigs in a village, or the number of doses of vaccine used), or as <i>real</i> numbers (fractional or decimal numbers, such as age, weight or temperature). |
| Integer numbers<br>Real numbers             |   |
| Categorical data<br>Qualitative information | • Categorical data uses numbers to describe what something is like ( <i>qualitative</i> information). Categorical numbers do not measure the amount of something, but are used to classify different categories. Categorical data can be divided into three types:  |
| Nominal data                                | • Nominal data (or ‘named’ data). Numbers are used to represent different categories, that are usually identified by their name. Numbers are used as codes, and the different categories have no natural order. For instance, species is often represented by a code, so that 1 = cattle, 2 = buffalo, 3 = pigs, 4 = chickens.  |
| Ordinal data                                | • Ordinal data (or ‘ordered’ data). Numbers represent different categories, but there is some natural order, so that 2 is greater than 1. For example, villages may be divided into small, medium and large, based on their cattle population. Ordinal codes could be used to classify the villages so that 1 = small, 2 = medium, and 3 = large.   |
| Dichotomous data                            | • Dichotomous data. Only two values are possible. This is the same as Yes/No data, but represented by two numbers (e.g. 0 = No and 1 = Yes).  |
| Dates                                       | • Dates. Dates are a special type of number data, used for data from questions like “When did you first notice the animal was sick?”.   |

### Data storage

The way data are stored in a computer is very similar to the way they are stored on paper. Let us first consider a paper storage system for survey data.

**Example:** A survey is conducted to estimate the seroprevalence of protective antibodies to rinderpest as part of a vaccination program monitoring scheme. The survey collects some data from a number of randomly selected herds, and collects blood specimens from randomly selected animals in each herd. The questions asked about each herd and the responses are stored in a file, with one sheet per herd. The data from the animals that were selected for sample collection are stored in a separate file, with one sheet per herd and one line per animal. These two files contain all the information collected during the survey.

In this example, the file for the herd data contains one sheet per herd. The data from the same question for each herd are recorded in the same place on each form. All the data on one form relate to one herd only. In the animal data file, each line in the table contains the data from one animal, and each column contains the same information for all the different animals in that herd.

In a computer database, information is stored as a *table* in very much the same way. A table is stored as a file in the computer's memory, and data from different things are kept in different files (for instance one file for the animal data, and one for the herd data). Each table is made up of columns and rows. Each row holds information about one item (for instance one herd in the herd file, or one animal in the animal file). Each column holds only one type of data, which is the answer to a particular question from each herd or animal. For instance in the herd file, the first column may store information on the name of the herd, the second has the number of animals, and so on.

In computer language, a row is called a *record*, and a column is called a *field*. When you create a new table for storing data from a survey, you need to say which fields (columns) you want to have, or in other words, what information you are going to store in the table. For each field, you also need to say what type of data will be stored there, text, numbers, yes/no, or dates. When you start putting data into the new table, you will add a new record (row) for each herd or animal that you enter.

A collection of one or more tables with related information is called a *database*.

## Data processing procedures

Once data collection has been completed, there are several steps before the data can be analysed. These are considered in more detail below.

Processing data for analysis

- Step 1:** Initial check for completeness and accuracy of data
- Step 2:** Data coding
- Step 3:** Creation of computer database
- Step 4:** Data entry
- Step 5:** Checking for errors and inconsistencies during data entry
- Step 6:** Recoding
- Step 7:** Converting data between different formats
- Step 8:** Analysis

## Initial check for completeness and accuracy of data

Before any work is done on the computer, the data record sheets need to be carefully checked for any missing data or mistakes. Preferably this should be done while the survey team is still in the village, so that the problem can be corrected. A quick examination of the sheets should show up any gaps where data have been missed. Finding other types of mistakes may require a closer look. For instance, a chicken's age may have been recorded as 12 years instead of 12 weeks, or a village stating that it has never had a disease outbreak has a record of the date of the last outbreak. If these types of problems are picked up early, then the question can be asked again, or the person filling out the form may remember the true response and be able to correct it. If this is no longer possible, the answer may have to be left blank, and treated as missing data. (See below).

## Data coding

Coding data is the process of converting complex data into a simpler form which is easier to manipulate. Computers are designed to work with numbers, so using numbers as codes makes the work easier. Using codes also makes data entry faster and more accurate and avoids inconsistencies

**Example:** A survey was carried out of 40 herds and 20 animals in each herd, with a total sample size of 800 animals. Data about each animal are being entered into the computer. Each herd is identified by its owner's name, and each animal must be identified with the herd that it came from. If the herd owner's name is used, then the whole name (which is sometimes very long) has to be typed for each of the 800 animals. It is very easy to make a mistake when doing this much typing. If a mistake is made, and the name of one herd is spelt in two different ways, then the computer no longer thinks that they are the same herd, but two different herds. This will cause problems for the analysis; for instance, the total number of herds is now 41 instead of 40. A better way to enter the data would be to use a numeric (nominal) code for the herd. The first herd is given the code of 1, the second is 2 and so on up to herd 40. It is much easier to type a short number than a long owner's name, and there are much less likely to be mistakes when entering the data.

### Data dictionary

Codes can be used for any type of data that has a number of different categories (categorical data). Herd, village, district or province names are some examples, but disease, species and season may also be converted to numeric codes. Before coding data a *data dictionary* has to be set up. This is a list of all the possible values on the data recording sheets, and the code to be used for those values. Sometimes this list will be easy to set up before the survey. For instance, if coding season in a tropical area, you might decide to use 1 = Wet, 2 = Cool, 3 = Hot.

In other cases, you may not know all the different responses until after the data have been collected. For instance when coding the responses to a question about which are the most important diseases of cattle that occur in the village. In this case, after the data collection is finished, the data recording sheets should be checked for all the different responses that were made, and a separate code assigned to each different disease that was mentioned.

### Missing data

Missing data is a problem that should be considered at this stage. When setting up codes, it is a good idea to have one code for missing data as well. Missing numbers need to be treated very carefully to avoid mistakes. When recording information on data recording sheets, missing data should either be left blank, or a dash “ – ” used. When entering data into the computer, some programs do not allow you to leave a field blank, and will insert a zero instead. This is a problem because there is a big difference between knowing that a villager has no ducks, and not knowing how many ducks they have at all. Depending on the program that is being used, it may be necessary to use a special code for missing data in number fields. For instance, the field storing information about the number of ducks may have the total number, or, if the data are missing, -99. During analysis, any villagers with a code of -99 for the duck population can be excluded.

### Epi Info

In Epi Info, if no data have been entered, this is automatically considered to be a missing value. During analysis it is ignored. However, if data are imported from a different program, they may be treated differently. Missing data that have been given a special code (such as -99) can be excluded from the analysis using the **select** command described on page 137. For example, **select ducks <> -99**.

## Creation of a table

Before entering data into the computer, a database program must be used to set up a table for the data. A separate table must be created for each different type of information. For instance, if both herd and animal data have been collected, two separate tables are needed. This is because each record (row) in the table stores data about one kind of thing only, either all animals, or all villages, but not a mixture of both.

#### Defining fields

A separate field is created for each different piece of data, and the type of data is specified. For an animal table in a seroprevalence survey, the fields may look like this:

Animal ID:	Number (integer)
Herd ID:	Number (integer)
Date of Visit:	Date
Age:	Number (real)
Sex:	Text (M or F)
Antibodies?	Yes / No

#### Field width

For some fields (such as text) it is also necessary to say how wide the field is or how much text will be stored there. In the above example, only one letter will ever be used for the sex field, so the width is one. For a field storing a village name, you may need room for 20 or 30 letters.

#### Data entry form

When the basic table has been set up, most database programs allow you to set up a *data entry form* as well. This is a screen on the computer with places to type each piece of data for a single record. Instead of showing all the data at once like a table, it shows only one record. When creating a data entry form, it should be in the same order and have the same appearance as the paper data recording sheet. This makes it much easier to find the data during data entry.

#### Data checks

One of the advantages of using a data entry form is that it is possible to set up *data checks* that check the data during data entry. Data checks can be of three types:

- Range checks • *Range checks* make sure that the value that is entered falls within a specified range. For instance, that the age of a buffalo is more than zero, but less than 30 years. If a number that falls outside this range is typed, the computer displays a warning.
- Allowed entries • Similar to a range check is a list of *allowed entries*. For instance, when entering the sex of an animal in a text field, it must be either M or F. Setting M and F as the only two allowed entries will prevent mistakes.
- Consistency checks • *Consistency checks* compare two or more pieces of information to check for inconsistencies. For example, data are recorded on the number of calves that a cow has had. When the number of calves is more than zero, the computer can check to make sure that the age is greater than 2 years, and that the animal is not a male. In either case, the computer will display a warning, and allow the user to correct the mistake.

**Epi Info**

In Epi Info, creating a new table and creating a data entry form are done at the same time. Use the following procedure:

Creating a new table

- Step 1:** Start Epi Info, and open **Eped**, the word processor. For detailed directions on how to use Eped, use the on-line manual.
- Step 2:** Design a data entry screen, which includes any text needed to describe fields or make data entry easier.
- Step 3:** Where data are to be entered into a field, insert field codes in the data entry screen. The main codes are:
  - (underline): text field (the number of underlines determines the size of the field)
  - # (hash): number field (the number of hashes determines the size of the field). Insert a decimal point if required, e.g. ###.##
  - <Y> yes/no field
  - <dd/mm/yy> date field (day/month/year format)
  - <mm/dd/yy> date field (month/day/year format)

**Example:** The questionnaire file for the survey shown above might look something like this:

Questionnaire file

```

Seroprevalence Survey Data Entry Screen

Animal ID:      #####
Herd ID:        #####
Date of Visit:  <dd/mm/yy>
Age:            ##.#
Sex:            —
Antibodies?    <Y>
```

- Step 4:** Save the file as a questionnaire file (.qes extension).
- Step 5:** Exit Eped, and start the **Enter** program from the Programs menu.

- Step 6:** When asked, type the name of the new data file that you want to create (usually the same name as the questionnaire file, but with a .rec extension).
- Step 7:** Select option 2 "Create new data file from a .QES file"
- Step 8:** Type the name of the questionnaire file you just created in Eped.
- Step 9:** Enter "Y" to indicate that everything is OK.

Epi Info will then create a new file according to the field definitions in the questionnaire file. Once the file is created you are ready to start entering data. However, you can also choose to set up some data entry rules and consistency checks within the file. Exit the **Enter** program, and use the following steps to set up checks:

#### Setting up data checks

- Step 1:** Select **Check** from the Program menu to start the Check program.
- Step 2:** Type the name of the newly created data file, or simply press enter and select it from the list.
- Step 3:** The data entry screen is displayed, ready to set up data entry checks.
- Step 4:** Specify different checks for all fields (see below)
- Step 5:** Press F10 to finish. Press 'Y' to save the changes to disk.

There is a range of different checks that can be performed on each field. Some of the important ones are listed below, but check the on-line manual for further details.

- The F1 and F2 keys will set the minimum and maximum values that may be entered (range checks). For many types of data, the minimum is 0 (no negative numbers allowed).
- Pressing the F3 key means that the data from the previous record are repeated in the following record. For fields that have the same information for many records in a row, this can save a lot of time. If the data are different, you can simply type the new data instead.
- The F4 key sets the rule that data must be entered in the field before the data can be saved. For critical data that must always be present (e.g. ID codes) this prevents errors.
- The F6 key can be used to set up a list of possible values (such as M and F for the sex field). No other entry will be accepted.

## Data entry

Entering data into the computer is a time consuming and boring task, and requires patience, accuracy and good skills on the keyboard. An experienced person can enter data quickly and very accurately, but it is always possible to make mistakes. Any mistakes made during data entry can lead to incorrect conclusions from the survey. There are several ways to try to avoid making mistakes.

#### Avoiding data entry errors

When entering a large amount of data, it is easy to lose concentration. Take a break every half hour or hour, and do something different. Alternatively, have two people work on the data entry and take turns.

Make sure the data recording sheets have been properly checked before data entry to make it easier. If there are any mistakes or numbers that are written unclearly, this makes it harder for data entry. Correct these problems first.

Using a database program with built-in range checking, allowed value checking and consistency checks will pick up mistakes when they are made, which makes them much easier to correct.

Double entry system

The best way to avoid mistakes during data entry is called the *double entry system*. Once all the data have been entered once, they are all entered a second time. During this second data entry, a special program compares the data being typed with the data that have already been stored in the computer's memory, and displays a warning if there are any differences. While it is easy to press the wrong key occasionally, it is very unlikely that the same wrong key would be pressed for the same figure during two data entry sessions. Ideally, the second entry should be done by a different person. An alternative double entry system uses two files that have been entered separately, and compares them for differences. Double entry is very good at avoiding mistakes, but has the problem that data entry requires twice as much work and takes twice as long.

### **Epi Info**

Entering data

In Epi Info, the **Enter** program is used for data entry. Select Enter from the Programs menu, and type the name of the data file (or press F9 to select from a list of files). Select 1 "Enter or Edit Data", and type "Y" for OK. A new data entry screen is displayed ready for entering data. A few tips might make the task easier.

- The number of the current record is shown in the bottom right corner of the screen. This helps you keep track of where you are up to.
- To edit data, use the F7 key to step back one record, or the F8 key to step forward one.
- To search for specific records, use Ctrl - F (hold down the Control key, and press F at the same time). You will then have to specify the information you want to search for (for instance a village ID number) and press F2 to do the search.
- Where a list of possible values has been set, you can press F9 to show the list, and use the arrow to select the value you want.

Double entry with the Epi Info  
Check program

There are two ways to use Epi Info for double entry system data checking. The first is to use the Enter program, but select 4 "Re-enter and verify records in the existing data file", instead of option 1. After entering the file name, you can enter data normally. However, if the data entered are different to those already in the file, a warning will appear.

Double entry with the Epi Info  
Validate program

The other way to check data is to use the Epi Info Validate program. Enter the data twice in two separate identical data files. Then run the Validate program to compare the data in the two files for any differences.

No matter how much care has been taken during data entry, there is still the chance of some mistakes. All data should be checked after data entry to pick up mistakes, as described below.

### **Saving and backing up**

Saving data to disk

While data are being entered, the computer usually stores the information in its RAM, the memory that works only when the computer is turned on. *Saving* the data means having the computer write the data to a file stored on a disk. Once the data are saved, the computer can be turned off without any risk of losing data. If the data have not been saved first, and the computer is turned off, all the data will disappear, and the work of data entry has to begin again. This can happen when a power failure occurs, or when there is a problem with the computer. Where the power supply is unreliable, these problems can be avoided by using an *uninterruptable power supply* (UPS), which has a battery to take over when the power fails. Even if a UPS is

Uninterruptable power supply  
(UPS)

connected to the computer, you should be careful to save the data onto a disk every five or 10 minutes. This will ensure that very few data are lost if anything goes wrong.

Backing up data

Data that have been written to a disk are much safer than data stored in temporary RAM, but there can still be problems. Occasionally, a disk can develop errors, so that the data cannot be read from it. Although this problem is rare, it can be very serious if there is only one copy of the data. Using regular *backups* will overcome this problem if it occurs. A backup is a second copy of the data which has been stored on another disk. Usually, the main copy of the data is stored on the hard disk inside the computer. Hard disks are very fast, and very reliable, but they do occasionally develop problems. Backing up the data to an external (floppy) disk, and keeping the disk in a safe place means that the data are not lost if there is a problem with the hard disk. You should backup your data during data entry at least once per day. After data entry is finished, the data should be backed up whenever any changes are made.

### **Epi Info**

In Epi Info, the data are saved to disk every time you change a record. The program asks **Write data to disk (Y/N<Esc>)?** every time. Answer Y to save the data.

To make a backup of the data onto a floppy disk, use your operating system to copy the file. In DOS, use the copy command (e.g. **copy results.rec a:**). When using MS Windows, use the File Manager or Windows Explorer to copy the files.

## **Checking for errors after data entry**

Once all the data have been entered they must be checked again. If this step is neglected, a lot of time analysing the data and producing reports can be wasted. Often, during analysis, an unusual result will be noticed which is because of an error in the data. When this error is fixed, all the analysis has to start again, because the data have changed. Rather than waste this time, it is much better to find the problems before analysis starts.

In contrast to checking for problems before data entry, checking after data entry is faster and easier, because we can use the computer to help us. The computer can search through all the data to find unusual values very quickly. Every field (column) should be checked separately, and then some fields may be checked together to assess consistency. Here are some useful techniques for using the computer to help check the data.

### **Epi Info**

In Epi Info, the data can be checked using the **Analysis** program. This is the same program that is used for general data analysis, and the same techniques described here are equally useful for that task. To start the Analysis program, start Epi Info, choose the Programs menu, and select Analysis. The Analysis screen is divided into two parts – the bottom part is for you to write commands to tell the computer what sort of analysis you want to do, and the top part reports the results of the analysis.

A few tips will help when using Analysis.

- To get a list of the different commands that can be used, press the F2 key. You can then highlight and select the command you want to use.
- To find out how to use a command, type the command and then press F1. This will bring up a help screen explaining the command and its use.
- Pressing F3 will bring up a list of the data fields in your table. You need to specify which field you want to work with in most of the commands.

Highlighting the field in the list and pressing enter is a fast way to insert the field name in the command. Sometimes you need to enter more than one field name. You can select many fields by pressing the + key, and then insert them in the command by pressing Enter.

- Use the F4 key to examine the data as a table. Pressing F4 again will show one record as a data entry form. Press Esc to return to analysis.
- As with all Epi Info programs, press the F10 key to close the program and return to the menu.

Before you can start working with a data file, you need to tell Epi Info which file you want. Use the **read** command and type Enter. You can search through the list for the file you want. Only Epi Info (.rec) files are shown. To use a file in dBASE format type **read \*.dbf**.

Counting records

The first thing to check is whether all the data have been entered, or whether some data have been entered twice. You can use the program to tell you the total number of records in the table, which should be the same as your sample size.

### **Epi Info**

Frequency tables

When you load a file in Epi Info, the total number of records is displayed at the top of the screen. Check that this matches your sample size.

Breaking the data into different groups to see the total in each group, and the number of different groups, is a very good way to check for mistakes. This is done by creating *frequency tables* (sometimes called one-way tabulations).

**Example:** A survey of villages was carried out in a single province. Some villages from each of the 6 districts were included. A code has been used to identify which district the village is in. By producing a frequency table of the district codes, it is possible to check that each of the 6 different district codes has been entered properly, and the correct number of villages is in each district. The frequency table might look like this:

DistrictID	Count
1	4
2	6
3	2
4	5
5	8
6	5
13	1
Total	31

In this example, seven different district codes are reported, with code 13 standing out as different to the rest. This indicates that a mistake has been made with the district code for one village, where 13 has been entered, probably instead of 03. If the total number of villages in the survey was 30, instead of 31, then one village has been entered twice. The total number of villages in each district could then be checked to find out which district has too many villages.

**Epi Info**

To produce frequency tables in Epi Info, use the **freq** command and specifying which field you want the program to use. The above table was produced by typing **freq districtid**, and pressing Enter.

Frequency tables are useful for identifying missing data as well. A frequency table of the sex of animals from a survey might look like this:

Sex	Count
F	124
M	32
Missing	4
Total	160

Frequency tables like this can be used for any text or categorical number field in which there is a limited number of possible alternatives (e.g. district codes, species, disease, sex, etc.). If frequency tables are used for other numbers (such as population) a long, unhelpful list of all the different population values is produced. For continuous values (integer or real numbers) there are two useful ways to check for errors.

Maximum and minimum

The first is to generate descriptive statistics for the field. In particular, the *maximum* and *minimum* values are useful.

**Example:** In a seroprevalence survey of pigs, the age (in years) and estimated weight (in kilograms) of each animal were recorded. Summary statistics were produced for these two fields, with the following results: Age: Minimum - 0.2 years, Maximum 100 years; Weight: Minimum - 3kg, Maximum - 2000 kg.

In both cases, the maximum is much larger than would be expected. They are probably an example of accidentally pressing the 0 key twice instead of once, and really should have been 10 years, and 200 kg. To check this, the original data recording sheets should be checked and the mistake fixed.

**Epi Info**

The **means** command can be used to produce a range of summary statistics. Type **means** and the field you want to use. The command usually produces a frequency table as well, which is not very useful. To produce just the results, type **means fieldname /N**. The /N means "no table". This is an example of the results displayed when examining a field containing village cattle populations:

```
====> means cattle /n
```

```
CATTLE
```

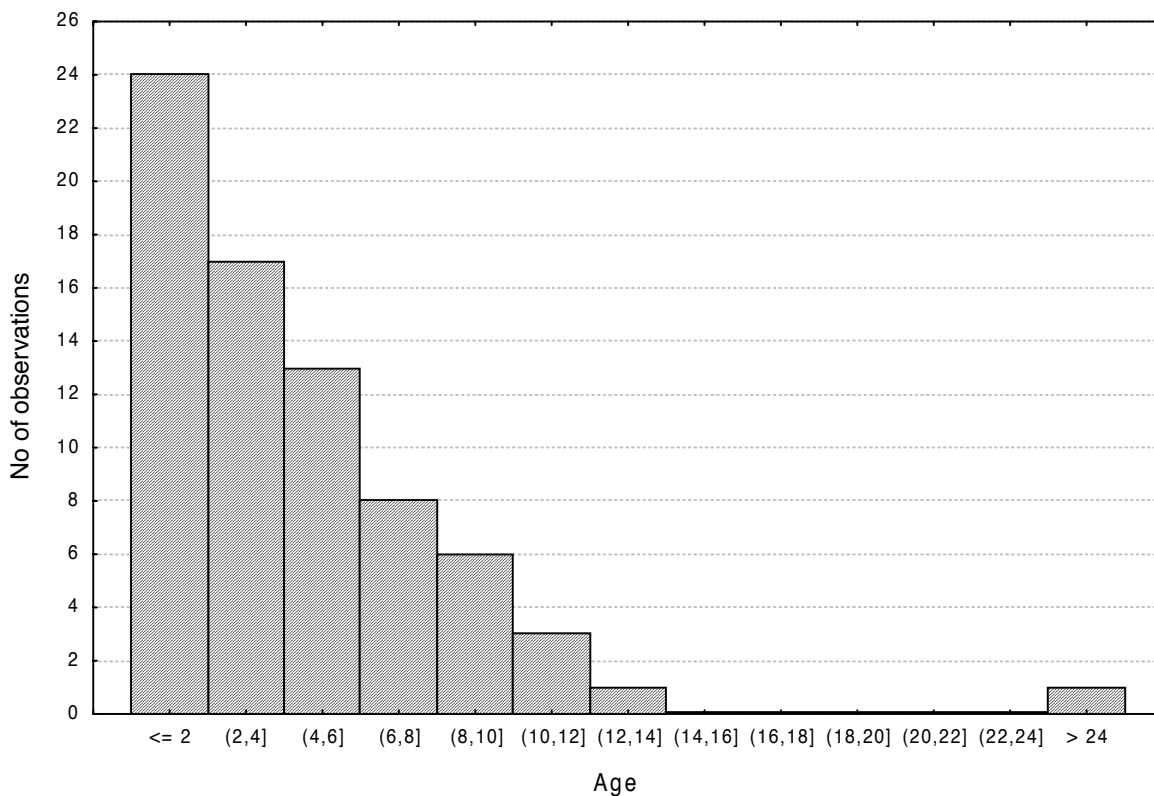
Total	Sum	Mean	Variance	Std Dev	Std Err
410	41510	101.244	11268.943	106.155	5.243
Minimum	25%ile	Median	75%ile	Maximum	Mode
0.000	24.000	66.500	138.000	621.000	0.000

```
Student's "t", testing whether mean differs from zero.
T statistic = 19.312, df = 409 p-value = 0.00000
```

Some of the information produced is not relevant, but Total (total number of records), Sum (sum of all the cattle in all villages), Mean (average cattle population), Minimum and Maximum provide valuable information.

The other way to examine this kind of data is to draw a histogram showing the distribution of values. A histogram is a graph which indicates the number of records that have a particular value, or a value in a particular range. Any unusual values will be on the far right or left of the graph, and can be easily seen.

**Example:** After correcting the problem with the age data identified in the example of the seroprevalence survey of pigs given above, a histogram was drawn, as shown below. The histogram shows the number of pigs that fall into different age categories, and indicates that mostly young pigs were included in the sample, with a few older pigs. There is one pig whose age is more than 24 years. This is probably a mistake, and the data should be checked.



Histograms can be used for both categorical data (e.g. codes, sex, species) and continuous data (age, population).

**Epi Info**

To show a histogram in Epi Info, use the **histogram** command, specifying the field to use. The above graph was produced by typing **histogram agegroup**, where agegroup identified which age group the animal belonged to<sup>8</sup>.

<sup>8</sup>In Epi Info, only categorical data can be used to draw a histogram. If your file contains data on the age of pigs, and you want to draw a histogram, you must first convert this into categorical data by grouping all the pigs with a similar age together into groups. To do this, create a new numeric field called AGEGROUP, using the command **define agegroup ##**. You then need to fill the new field with a code that defines the age group of the pig. In the histogram, pigs have been categorised into two year groups. Use the command: **agegroup = 2 \* round(age / 2)** to create the categorical age group codes, grouped into 2 year brackets. To group into 5 year brackets, replace both 2's with 5.

Selecting subgroups

In addition to examining one field at a time, it is possible to check two (or more) fields at the same time for consistency. One way of doing this is to select a group of records from the data that match a certain definition.

**Example:** A survey of cattle has recorded sex, and the number of calves born. A group of records was selected that was identified as Male, but didn't have 0 as the number of calves born. The definition for the group was: (Sex = M) and (Calves > 0). The number of records in the group was then counted to reveal that 3 records matched the definition. These three records were checked with the original to correct the mistake.

**Epi Info**

To select a group of records in Epi Info, use the **select** command, specifying which records you want to use. After you use the select command, any analysis you perform will work only with the selected subgroup. To turn off the selection and work with all the records, use the **select** command without specifying any records. Using several different **select** commands selects only those records that match each of the commands. For example:

```
select sex = M           only male animals are selected
select calves > 0       males and animals with calves are selected
means calves /n         summary statistics (including a count of the number of
                        records) are produced.
select                  turn off the selection - all records are active
```

Cross tabulations

Two-way tables, or *cross tabulations* can be used to check two categorical variables as well. Using the same example to compare sex and number of calves, a two-way table may look like this:

	Calves					
Sex	1	2	3	4	5	6
F	122	64	35	12	0	1
M	0	1	0	2	0	0

This table shows the number of records that match two different criteria. For instance, there are 122 records of female animals with 1 calf, and 64 females with 2 calves. The three records with errors can easily be seen.

**Epi Info**

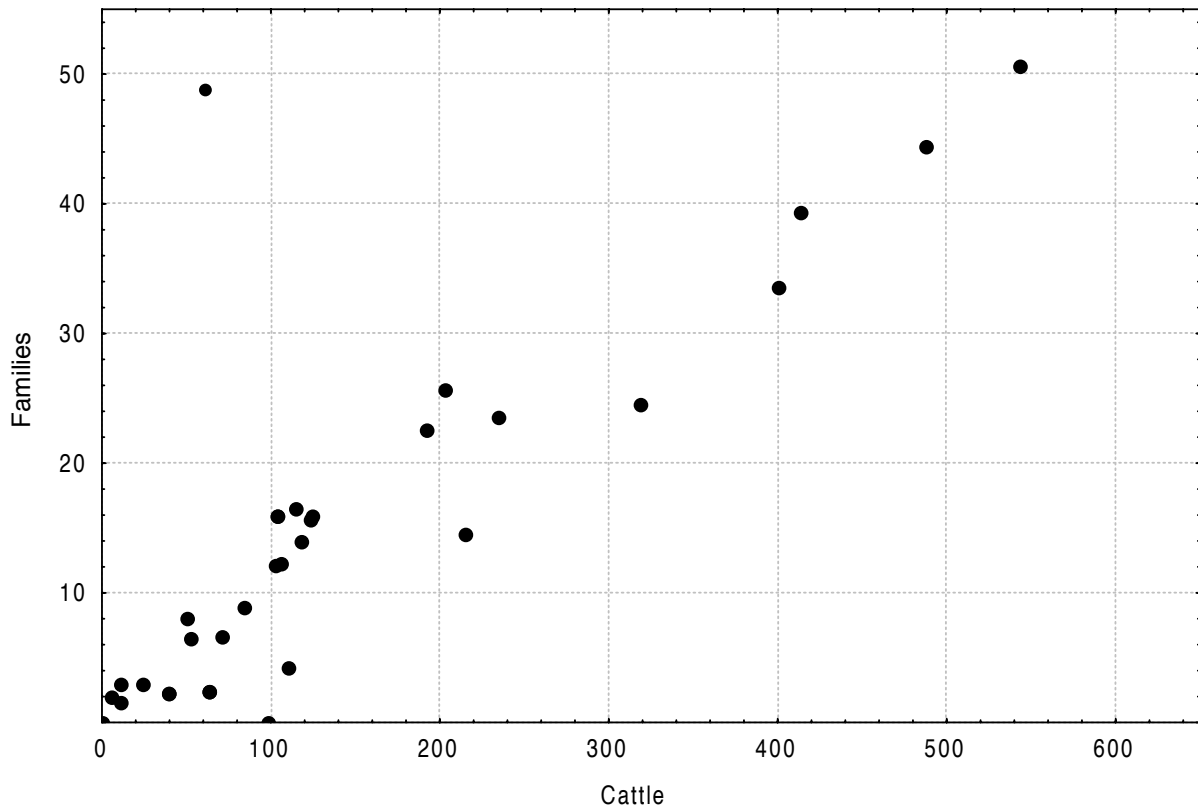
Use the **tables** command to produce two-way tables, specifying the two data fields to include. The above table was created with the command: **tables sex calves**. Note that the first field you specify is shown in the rows, of the table, and the second in the columns.

Scatter plot

When two different types of continuous data are to be compared graphically, you can use a *scatter plot*. A scatter plot draws a point showing the value of one of the variables on the x axis, and the other on the y-axis.

**Example:** Village data have been collected on the number of families raising cattle and the total number of cattle in the villages. These two groups of data were displayed on a scatter plot as shown below. Most of the points lie in a line close to the centre of the graph, indicating that villages with a small number of families with cattle, usually have a small total number of cattle, while villages with many families have many cattle.

There is one point at the top left of the graph that is well away from the rest. This point indicates a village that seems to have many families that own cattle, but only a small total number of cattle. While this is possible (every family has only one or two animals), it is clearly unusual, and may be a data entry mistake. The original data should be checked.



### Epi Info

To draw a scatter graph in Epi Info, use the **scatter** command, specifying which two fields you want to use. The graph above was produced with the command: **scatter cattle families**. The first field is shown on the x axis.

## Recoding

Data are not always in the best form for analysis when they have been entered. Recoding is the process of changing data from one form to another, to make it easier to analyse. Recoding after data entry is made much easier, as the computer can do all the calculations for us.

**Example:** In an incidence rate survey, the date of the last outbreak of Newcastle disease amongst village chickens was recorded, along with the date of the visit to the village. For analysis, we need to know how long ago the last outbreak was, not the date. This is calculated by subtracting the date of the outbreak from the date of the visit.

**Epi Info**

Recoding data is done by telling the computer how you want to change the data. It usually involves creating a new field. The computer calculates the values for this new field based on other data in the table.

In Epi Info, you must first create a temporary field to store the new coded value. Use the define command, specifying the name for the field and what type of data will be stored.

For instance **define herdname \_\_\_\_\_** creates a new field with the name "herdname". This is followed by 10 underlines (\_) which indicate that it is a text field, with room for 10 letters. Number fields are represented by hash marks (#) so the command **define weight ###.##** would create a new field called weight, for storing numbers. The numbers would have room for 3 digits and 2 decimal places. A date field is defined as **define visitdate <dd/mm/yy>**.

Once the new data field has been created, you can tell Epi Info what data you want to be put in that field. Some examples of commands include:

<pre>define time #### time = visitdate - eventdate</pre>	<pre>Create a new field The time between two dates (in days) is calculated and stored in 'time'.</pre>
<pre>define agemonths ### agemonths = ageyears * 12</pre>	<pre>Create a new field The age, expressed in years, is converted to the age in months.</pre>
<pre>define totalpop #### totalpop = cattle + buffalo</pre>	<pre>Create a new field The population of both cattle and buffalo is calculated and stored in the 'totalpop' field.</pre>

## Linking data

Sometimes all the data are not in a single table, but in two or more tables. Before analysis, the tables have to be *linked*.

**Example:** A seroprevalence survey was carried out in which 250 blood specimens were collected from pigs. During specimen collection, data on the sex, age, and vaccination history of each pig were collected. At the end of the survey, these data were entered into a computer database. The blood specimens were sent to the laboratory for analysis. The laboratory uses a computerised recording system, and was able to send all the results from the blood tests on a computer disk. This means that a lot of time was saved by not having to retype the blood results, and typing mistakes were also avoided. However, the laboratory tested the blood specimens in a different order to that used on the data recording sheets. The two tables need to be linked.

Key field

To link two tables, there must be one piece of data which is the same in both tables, called the *key field*. The computer uses this field to know which piece of data from one table belongs to which piece in the other table. The data recording sheet identified each specimen by a tube number. The laboratory also identified each of the test results by the tube number. The tube number can therefore be used as the key field to link the two tables.

**Epi Info**

The actual process of linking depends on the database program being used. In Epi Info, link two tables using the **relate** command, and specifying the name of the key field, and the name of the file to join. To join to files:

- Step 1:** Open the first file in analysis, using the **read** command.
- Step 2:** Make sure there is a key field with the same field name in both the open file and the file you want to link.
- Step 3:** Use the **relate** command to link the second file.

**Example:** Using the above example, there are two data files, one called **animals.rec** and one called **results.rec**. Both files have a field called **tubeID**. After first opening the animals file, the two files are linked using the command **relate tubeid results**.

Making a link permanent

Once tables are linked, the data can be analysed. However, the link is only temporary, and when analysis is finished, there are still two separate tables. To make the link permanent, save the linked table to a new table. In Epi Info this is done in two steps: first define a new data file name (the **route** command), and then write the information to the file (**write recfile** command). For example, to save the two linked files to a new single file called **alldata.rec**, use the command **route alldata.rec**, and then the command **write recfile**.

## Converting data between different formats

If using Epi Info for data management, the same program can be used for data entry, checking, coding, linking and most of the analysis. However, there are times when you may need to use a different program for analysis. For example, there may be a special type of analysis that is required for particular survey types, such as those described in this book. The programs to carry out this analysis are included on the disk, but are not part of the Epi Info program. Alternatively, you may prefer to use a different database program to manage your data, and then need to use a specialised statistical program to analyse it. Sometimes data are provided by somebody else and may have been prepared using a different program. In all these cases, the data must be converted into a form that can be used by the different program.

Format

The *format* of a data file is the way it is stored on the disk. Different programs use different formats, as they store data in different ways. To use data in a different program you need to change the data format to a format that the analysis program can use. Commonly used formats include dBASE, Paradox, and ASCII.

Exporting data  
Importing data

There are many different data formats, but fortunately there are a few standard formats that are used to move data between different programs. Examples are dBASE files (with a .dbf extension) or Paradox files (.db extension). Many programs are able to both read and write data in these formats. Writing data to a new format is called *exporting*, and is usually accessed through the Export or Save As menu of a program. Reading data from a different format is called *importing* and can be accessed through the Import or Open menu option of most programs.

Epi Info has two special programs for converting data between different formats, both of which can be accessed from the Programs menu. The import program reads data from four different formats, including dBASE, and saves it as an Epi Info format (.rec) file. The export program is able to convert Epi Info data to 16 different formats (including dBASE). As mentioned earlier, Epi Info is able to use dBASE files for analysis without conversion.

All the Survey Toolbox programs can read and save data in either dBASE or Paradox formats.

**Epi Info**



## **Analysis**

Finally, when all the data processing is complete, it is ready for analysis. Analysis of simple data can be done using the same tools and methods described above for checking data: frequency tables, counts, two-way tables (cross tabulations), simple descriptive statistics (maximum, minimum, average), and graphs (histograms, scatter graphs).

More complex data analysis is necessary for complex survey designs. The methods and programs for the analysis of data from the surveys described in this book are explained in Part II.

# A I

## l o l o

ar r n r n r arr n r  
r n r r a n r ; n n ra r ; an r  
n ra r r a n a r  
r n r r n a a r ar a r n  
n a a n arr r r r n

Survey procedures

S r n a n n a an an r  
S n ar a n  
S r r n  
S r ra a n  
S a a a  
S an a  
S ran r a  
S n a r  
S a  
S arr r  
S n r a n r n r an ran a  
S r n r a r ana  
S n n a ra r  
S a a r n an a ra  
S n r r a a an a ra r r n a r  
S a a r a r n a a n r  
S na a a a a  
S r a a

ar r n

n n ra r ar a  
r

r n a n n a  
n ar a n  
r r n  
a a a  
an a  
ra n r a  
n a r  
na a a a a a  
r a a

**O**

r a r a an ra na an a  
a n n an r nn n an r n  
an r an ran r nr nnn a a r  
a n r a n r n ra r r n an  
a rra a a a na n r ra r n r  
ra r n n r nr r r an n n r  
n a a r n n nr  
n ar a ar n n n a ara  
an r n r n a an a r  
an an r n r n an n n  
r n an r n a na n r ra r n  
r ra r n n r r n an a n a n r  
an r r an ar n n a  
r a n n a n r an ra  
n a r a an ar ar a  
an a ar a r a r n n a a  
r r  
r r n n a nan r n n a a ra an a  
an r n a a ra an r n r ar  
ra r n a an r

**O I O**

Target population

n a r r ar ra r n a n a ar  
n r n r u a n a a an r  
na n r an r n r n nr n

Source population

ar a n a n an a n nr  
ur u a n a a ra a r  
an r r na a r a n n n na  
n n ar an a a an a n r  
a n a a a ar a n a  
r a ar ra a a r n an a  
a r an a r r a a a ar r  
r a n  
ran r r a r r an a  
n r n a r a n n n ar a n ar  
r n a ra n n r n

oo

a a n r an r n r n n a a  
 r n n n a a n a a na n r ra n  
 a n an ar a a r a n r n ar n  
 a n n ra a a ar ar a n r n nar n r nr  
 a r n ra r r a a r ra ar r  
 r an n n ra r r a n r a ra  
 n r r n n r an n  
 r r r arr n r an n a r  
 r a n an n n ra a n n  
 a r ra

I

r ar n a r n n an n n r r  
 a ran a n a n an r n n  
 r ar ann n a r ar r a r r  
 r n r a r n n n an r r ar  
 r ar r n an n n r  
 a r n r nr r a r n r a a a n  
 an r n n r a n r a a a n r a a n a  
 ar an n n a r a r n  
 n a a a n ar r ana r r n  
 a ar aran a n ara n n a n r  
 r n an r n n r a r r r  
 n r a n r a a a n a r n n a r

aran a a r ara n r n a n r  
 r n r n n a r r a aran an  
 r an an a a a n ar r r a n  
 a r a a a a aran a a n a a r a  
 a a aran

n r r n a n a n a a  
 n a r a n a n an r an n n  
 r ar n r n a a n a ra a an a  
 n a n r a a n a r  
 n a n an an a r an r a  
 a n an r n a an a ar n n  
 a a n n r n a a r a a r ar an  
 an an r

aran an ran a r n a a n a na a n  
 a aran an a ar r ar a a ra ra  
 r a n n n a an a ar r na r a  
 ra a n aran ran a ar n a a  
 an a r n a n ar r  
 n aran a n r

ar r n

o o

r n a r n a a a  
 a a r n n n r a a r  
 n n n r a a ar n r r a r r a  
 n r an anarr n n n r a n a a ar r r a  
 r a r nanarr ran a r r a  
 n n a a r a r ar a a n  
 a ran r n n n r a n n an a ar  
 r a r a a n n r a a r a n  
 na ra r n n r  
 n n n a a n n a an  
 ar n a a n a r a r r n a  
 narr r n n n r a r r a ar r a  
 r n n r a a n n n r a r n an  
 n r n a r a a

l l o

a n n ann n r r a r ar r a r  
 r n r an r n a n n ar arr  
 ar a n r a a n n a a a  
 a a a n ar r r r n  
 n n a a a a n ar n r a a  
 n ar na a n

Planning checklist

- a n a an ar
- ann a r n
- ann r ar n an r n
- ann a
- n a
- n n a a
- ann r ran n r a r n r n
- a r
- ann a ar a
- r ar n a r a r r ana n
- r ar n a ar r n
- ann n r r a n r
- rann a
- n n r n a ar r n an n r a
- ann an n r r a a ana n r r n
- rann a r r a a n r r a a

ran a a r r n r r  
 rr a n n r r a n r a n n r a  
 n n an a n a an r a r n r r r r  
 rann r ra n a n r ara n ran r  
 n a a rann r r an ar n n r  
 ran r n an an rann a

# lo

r a a r a arr r r a r  
r ar r ran r r a n

n r a n n r a  
n an r n n r r n  
n r a n n a n a an a a a  
r a r r a r a an  
n r ar an n r n r r n ar a a n  
r n

# A I

r r r a a ana n a  
r r n r n an ana  
r ar n n a an n n r n r  
ar ar n n r an a r  
ar r ana r n a r r r r  
n a a n n na r r n ra a ana r r  
r ar r an n r an n r  
n ra a ana r r r n a n  
n r arr ran ana a a n ar r  
n n n n an a an a ar na r ra

# O

a a r n a a r n ra  
r an ra n an r n a  
r a n a r an a a r r a  
an r n r an a n n n ra  
r r n a ar a r a  
r r a n n ra  
a r r r n an r nar a ar a r a r n n  
a r ar r a n a n

r n a r an n r n n  
a a ana an r ara n r r r a a a r r a a  
a r r a n r n a ar r  
n r a n n n a a n n n r r  
a r r r a na a r r n  
n r r a a ana a r a r a ar  
an r an r r ana n ana n  
a a an r n r r n n a r a  
an r n n n an a an r n  
n r ran n r a n n r r r  
n r n a a a a n a a r r  
a r ar a an n n a n r an  
ana n n r an a n ar n n an  
n r r r a r an a  
n r a n r n na a a a an  
n r an a r an n n r na a r r

ar r n

r a r a a a a a r r n r a a a  
a a ar ar ar n r a r a a  
n r a n r r a n an  
r ar a n n r a n a a an a  
r nar a a a na na r nar a a r r r  
a r a an r an rr rn  
n r a n a n r a r an a  
a r a a a a n n r a a a  
n an a a a n n r  
a n r n r n rna na  
rn nr n n n r a n n r na n  
ra r na a r a a nr r a r a n  
n r a n n n rna na an a a r an a n a  
r na n rna na an an  
r r r an a n n a n  
r n n rna na rna ar  
r n r a n r a n r a arr an r  
an r ann n ar r r r an an arr r  
r n r a a arr a r an  
n rna na r n an ar r a r n rna na  
r a n a nr r nar a r n n an a  
n n a n n rna na ra

# 7

## Prevalence Surveys

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys  
    Conducting a survey  
    Choosing the right design  
    Sample size  
    Stratification  
    First-stage sampling - the village or herd  
    Second-stage sampling - the animal  
    Data analysis

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



## Prevalence surveys

Prevalence surveys aim to estimate the proportion of the population that had a particular disease or status, at a single point in time (see page 26). Prevalence surveys are the most commonly used way to gather information in livestock disease surveillance programs. This chapter describes a series of survey designs developed especially for developing countries. They are able to gather unbiased, reliable data, as quickly and as inexpensively as possible.

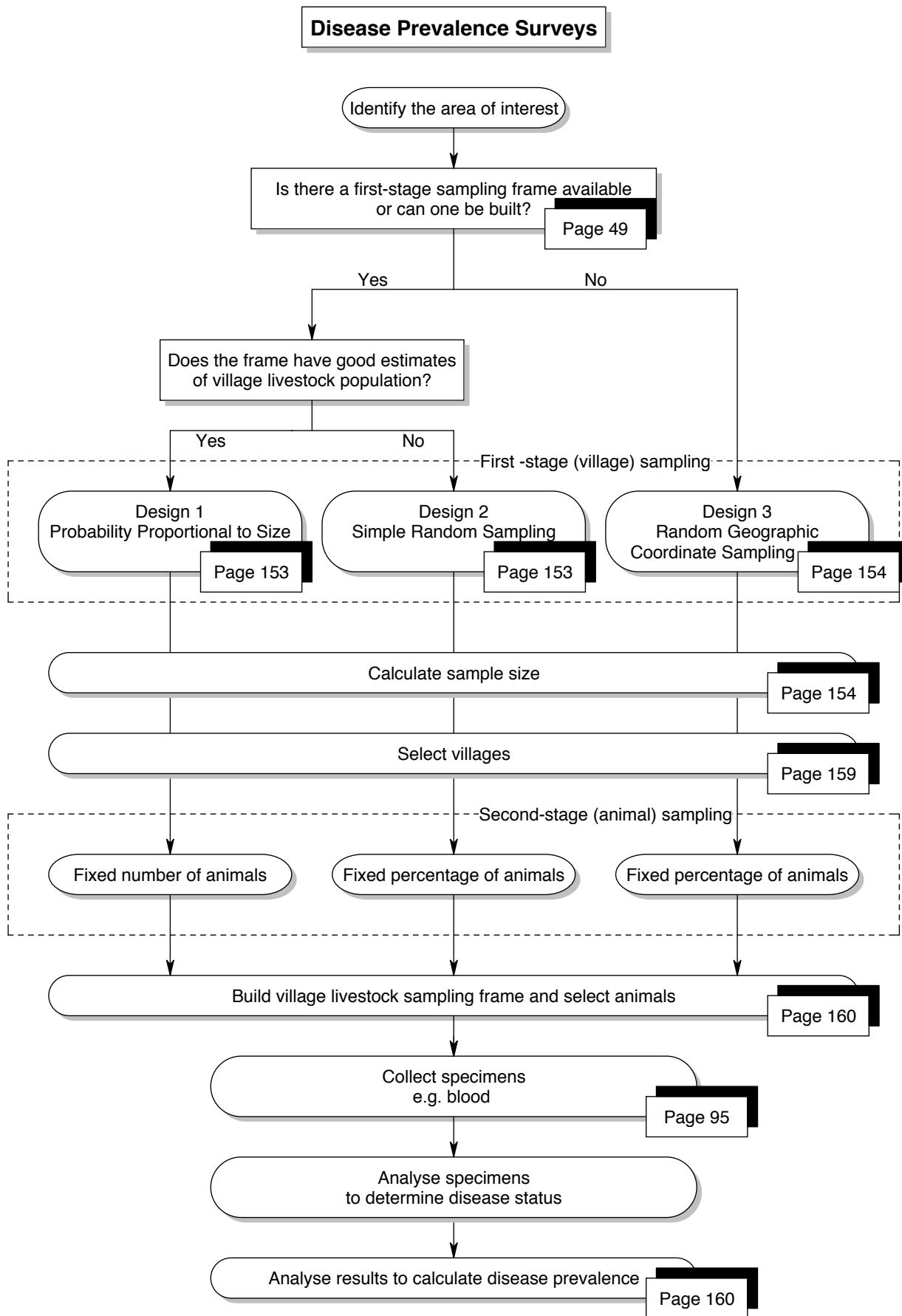
Prevalence surveys may be used to assess disease priorities and develop control strategies (determine how much disease there is in a population, where the disease is occurring), or to monitor the progress of a control program (e.g. determine the proportion of animals with antibodies due to vaccination). Both prevalence and incidence rate surveys measure the amount of disease in a population, but in different ways. The difference between the two measures is discussed on page 29, to help decide which one is best for a particular situation. Chapter 8 describes how to conduct an incidence rate survey. Surveys to demonstrate freedom from disease are similar to prevalence surveys in that they aim to identify diseased animals, but the design and analysis are quite different. These surveys are described in Chapter 9. While the three different survey types are dealt with separately, it is possible to collect information to estimate two or three of these measures during the one survey.

## Two-stage sampling

Surveys of large areas (national, province, state or district surveys) which measure prevalence at the animal level are difficult because of the lack of a sampling frame. To use random selection (and ensure reliable results) requires a sampling frame which includes every animal in the entire study area (see page 49). Building an accurate sampling frame of all animals in a large area is usually impossible. *Two-stage sampling* (page 64) avoids this problem by breaking the sampling into two steps. First, groups of animals are selected randomly (usually villages or herds). At the first stage, a sampling frame listing all the villages or herds (the first-stage units of interest) in an area is all that is required. Once the groups are chosen, each village or herd is visited, and a sampling frame of the animals in the group is constructed, and used to select animals (the second-stage units of interest) for the sample.

The strength of two-stage sampling is that animals need to be listed only for a small number of herds or villages, rather than the whole population. In addition, the field work is easier, as the survey team has to visit only a relatively small number of villages. If simple random sampling was used, there may well be one or two animals from very many villages, which would require a lot more travel. The weakness of two-stage sampling is that the survey design and analysis are more complex. The Survey Toolbox provides programs that make these jobs much easier.

This chapter is a guide to conducting two-stage prevalence surveys for livestock diseases as part of an active surveillance program in developing countries.



## Conducting a survey

There are 20 main steps in running a two-stage prevalence survey, as shown below. This description is based on a survey of village livestock, in which specimens are collected for laboratory analysis (e.g. blood samples) to determine the disease or antibody status of the animals. Some of the procedures will need to be modified slightly (simplified) if 1) no specimens are collected, and only clinical examination is used to determine the disease status of animals, or 2) herds rather than villages are selected in the first stage, in which case it may be easier to select animals randomly.

- Step 1:** Determine what question is being asked and how best to answer it.
- Step 2:** Identify the target population.
- Step 3:** Choose the right survey design. There are three different designs (PPS, SRS, and RGCS), based mainly on the way herds or villages are selected at the first stage of sampling. The choice depends on what sampling frame is available.
- Step 4:** Calculate the best sample size. A computer program is provided to help with this, but some knowledge of the disease and population is also needed.
- Step 5:** Decide if the survey is to use stratification, and if so, what basis will be used.
- Step 6:** Plan field activities, decide on interview questions, prepare data collection sheets, transport, restraint equipment, specimen collection and processing equipment.
- Step 7:** Train survey teams.
- Step 8:** Select the first-stage sample (herds or villages) using random sampling.
- Step 9:** Visit selected herds or villages.
- Step 10:** Conduct a village interview, to build an animal sampling frame for the village, and ask other questions.
- Step 11:** Select the second-stage sample (animals) using random sampling.
- Step 12:** Visit livestock owners, and identify selected animals.
- Step 13:** Restrain animals and collect specimens.
- Step 14:** Process specimens ready for analysis.
- Step 15:** Send the specimens to the laboratory.
- Step 16:** Check the data for completeness and accuracy.
- Step 17:** Enter the survey data and laboratory results into a computer.
- Step 18:** Check the data for mistakes during data entry.
- Step 19:** Analyse the data to estimate the prevalence.
- Step 20:** Report the data, providing feedback to livestock owners, local veterinary staff, national veterinary authorities, and perhaps international publications or organisations.

Some of the key steps are described in detail below.

### Step 3: Choosing the right design

There are three different survey designs for two-stage prevalence surveys. The choice of the best design to use depends on what sort of sampling frame is available for first-stage sampling – a sampling frame with population data, a sampling frame

without population data, or no sampling frame at all (see page 49 for a full discussion of sampling frames). The type of sampling frame determines how herds or villages are chosen at the first stage, and how animals are chosen at the second stage.

### Design 1 (PPS)

Probability proportional to size

In the best situation, a complete sampling frame is available, listing all herds or villages and including *reliable livestock population data*. This may come from data maintained by the veterinary services, who regularly update livestock population figures, or else a recent agricultural census. When this information is available, probability proportional to size (PPS) sampling may be used.

In Design 1, villages or herds are chosen at the first stage so that the chance of selecting one with a larger population is greater than one with a smaller population. When selecting animals at the second stage, a *fixed number* of animals is chosen from selected villages using simple random sampling.

**Example:** Design 1 (PPS) was used for a prevalence survey of village pigs. An agricultural census was carried out 1 month earlier, and the data were used as a sampling frame. Forty villages were chosen with probability proportional to pig population. Each of the villages was then visited, and a village interview of pig owners used to build a sampling frame for the village. In each village 15 pigs were selected by simple random sampling from the sampling frame.

This survey design is the most efficient, as it is able to make more accurate estimates for a given sample size than the other designs. It is also easier for the survey teams when doing the field work. Unfortunately, while a complete sampling frame might be available, it is quite uncommon to have complete up-to-date data on village or herd livestock populations for the relevant species. If the data are only a few months old, they will already be incorrect. If there are only small changes in the population, this doesn't matter too much, but if there have been large changes in some villages or herds, the population data can no longer be considered reliable, and Design 2 should be used.

### Design 2 (SRS)

Simple random sampling

When a good sampling frame, containing all the herds or villages in the area is available, but there are no reliable population data, simple random sampling (SRS) can be used at the first stage.

In Design 2, every village or herd has the same chance of being selected. At the second stage, a *fixed proportion* of animals are selected from the population using simple random sampling, instead of a fixed number as is used in Design 1.

**Example:** A survey of village chickens uses a Statistics Office list of all villages in a province as the sampling frame. No figures are available for the chicken population in the different villages. At the first stage, a sample of 40 villages is selected by simple random sampling. Each of these villages is visited, and the chicken owners are gathered for a village interview. A sampling frame is constructed, and in each village 5% of the population is selected at random for the sample. In a village with 252 chickens, 13 chickens are selected, in another village with 689 chickens, 34 chickens are selected.

Design 2 is reasonably efficient, but not quite as good as design 1. Because the survey teams don't know how many animals there will be in the villages or herds before they visit, they don't know how many animals will need to be examined or specimens collected. In a large village, there will be a lot of work, in a small village, not much at all. This makes planning the field work slightly more difficult. However, most of the time a sampling frame is available, and this is the survey design that should be used.

### Design 3 (RGCS)

Random geographic coordinate  
sampling

In the worst case, there is no sampling frame for herds or villages available at all. This is usually the case for nomadic herds, or when government structures and records have broken down due to war or other disasters. The only way to select a random sample of herds or villages at the first stage is to use random geographic coordinate sampling (RGCS).

In design 3, RGCS is used to select herds or villages (see page 65). At the second stage, a *fixed proportion* of the village population is sampled, just as with design 2.

The statistical efficiency of this survey design is similar to that of design 2, but the field work is much more difficult. This is because a lot of field work is needed before the actual survey, to select the villages or herds. For this reason, design 3 should be used only if necessary. Usually it will be possible to find a reliable sampling frame.

## Step 4: Sample size

The sample size is calculated using the computer program described below. For a two-stage sampling survey, the sample size is made up of the number of herds or villages to sample at the first stage, and the number or proportion of animals to sample at the second. It is possible to select fewer villages and more animals, and still get results of the same accuracy. This makes two-stage sampling very flexible, and allows the survey design to be adjusted to achieve results of a specific level of accuracy, but at a minimum cost. Calculating the best combination of first- and second-stage sample sizes requires a few different pieces of information, described below. When a survey is being carried out for the first time in a particular area, some of these numbers may not be known, and estimates have to be used. However, when a survey is used as an ongoing part of a surveillance system, detailed information is available from previous surveys, and very accurate calculations of the minimum cost sample sizes can be made.

### Survey costs

The costs of the different parts of the survey need to be known – the cost per animal, and the cost per herd or village. It is the ratio of these costs that determines the least expensive combination of first- and second-stage sample sizes.

The per-animal costs are mainly made up of costs for laboratory testing, and equipment, such as blood tubes, needles, serum tubes, etc. They may also include a cost for the salaries of the field staff, which is based on how much time it takes to examine or collect specimens from each animal. Per-village costs are usually made up of field staff salary and transport costs.

These costs are summarised in the table below. Other costs that do not vary with the number of animals or the number of herds or villages (e.g. the cost of obtaining a village sampling frame) are not included in the calculation.

Per-animal costs	Per-village costs
Blood tubes	Fuel
Serum tubes	Vehicle costs
Laboratory tests	RGCS costs
Staff salaries	Staff salaries

When conducting a survey in an area for the first time, it is useful to keep accurate records of the costs involved. These figures can be useful in planning future surveys. When no previous figures are available, the costs need to be estimated.

### **Variance**

In two-stage sampling, there are two populations that are being sampled, the herds or villages, and the animals. Each of these two populations has its own variance (see page 145). The amount of difference between different herds is known as the “between-herd variance”. The spread of difference between individual animals within the same herd is called the “within-herd variance”. When calculating the sample size for a two-stage survey, both these variances are taken into account by the computer program. These values are very hard to estimate, so either values from a previous survey need to be used, or estimates based on similar surveys in other parts of the world.

For example, seroprevalence surveys of foot and mouth disease antibodies of cattle and buffalo conducted in Southeast Asia have yielded within-herd variance estimates in the range 0.15 to 0.22, and between-herd estimates between 0.03 and 0.08. If no other data are available, these figures can be used as a starting point for initial estimates.

### **Population size**

The total size of the population is needed for some sample size calculations (depending on the survey design). Where full population data are available for every village, this is not a problem. However, where no data exist, the total population must be estimated. Fortunately, it doesn't matter too much if this estimate is not perfect. There are usually some records available for the population in an area.

### **Estimated prevalence**

One of the most difficult things to understand about calculating sample sizes for prevalence surveys, is that you need to know approximately what the prevalence is before you do the survey. For surveys held as a regular part of an ongoing surveillance program, earlier prevalence estimates will allow good estimates to be made. However, for the first survey in an area, guesswork will be needed.

### **Area of study area and selection radius**

For random geographic coordinate sampling (Design 3), an estimate of the total area of the study area is needed. These figures are often available through the National Statistics office. If digital maps of the study area and geographical information system (GIS) software are available, the size of the study area can be calculated easily. If only paper maps are available, you can still work out what the approximate area is, by drawing a grid of squares over the study area, and counting them.

An estimate of the anticipated selection radius is also needed.

Relative error

**Precision**

Precision is usually measured as the width of the confidence interval. A fixed width may be used, or, for very low or high prevalences, the *relative error* may be better. This is because as the prevalence gets smaller, we often want to measure it more precisely.

**Example:** Using a fixed width confidence interval of  $\pm 5\%$ , a survey resulting in a prevalence estimate of 50% would have a confidence interval of 45% - 55%. This is probably precise enough for most purposes because the difference between 45% and 55% is unlikely to be very important. If the prevalence was 5%, the confidence interval would be 0% to 10%. The difference between 0% and 10% is probably quite important, so we would often want to measure the value more precisely if the prevalence is low. If we used relative error, the confidence interval for a prevalence of 50% may be 45% to 55%, but the confidence interval for a prevalence of 5% may be 3% to 7%.

The relative error is a measure of the width of the confidence interval as a proportion of the prevalence, so the smaller the prevalence, the narrower the confidence interval.

For fixed-width confidence intervals, a value of  $\pm 5\%$  or  $\pm 10\%$  is commonly used. If a smaller value is used, the sample size will increase in size dramatically. A relative error of 0.1 will produce a confidence interval of about  $\pm 10\%$  if the prevalence is about 50%, but if the prevalence is 10%, the confidence interval will be about  $\pm 4\%$

**Confidence level**

The confidence level determines how confident we are that the true value lies within the confidence interval. By convention, a confidence level of 95% is used most of the time. This means that in one case out of 20, the true value may lie outside the confidence interval.

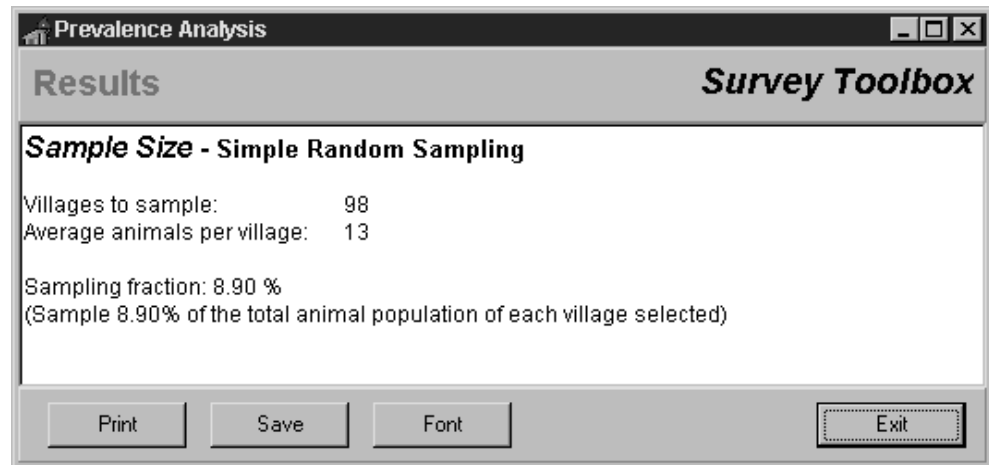
**Calculating the sample size**

The formulae for calculating the sample size for the three different survey designs are very complex, and can normally be calculated only by a trained statistician. In order to enable non-statisticians can do the calculations, the formulae have been incorporated into the **Prevalence Analysis** program, included on the CD. To start the program use the Windows **Start** menu, select Programs, Survey Toolbox, and Sample Size. To calculate the sample size required for a two-stage prevalence survey:

- Step 1:** Click on the Sample Size Calculation tab at the top of the window.
- Step 2:** Select the survey design to be used. In the First Stage Sampling Scheme box, select either Design 1 (Population proportional to size sampling - PPS), Design 2 (Simple random sampling - SRS), or Design 3 (Random geographic coordinate sampling - RGCS). See *Choosing the Right Design* on page 152 to help you decide.
- Step 3:** If you are unsure, click on the **Which one?** button for help deciding which design to use.



- Step 4:** Look at the Second Stage Sampling Scheme for advice on how to select animals at the second stage.
- Step 5:** In the Parameters box, enter all the parameters required, as described above. You must enter estimates for all the parameters. You need to decide on the required accuracy and confidence level yourself.
- Step 6:** If you have the results of a previous survey, you can use these to enter the first- and second-stage costs (cost per herd and cost per animal).
- Step 7:** If you have a data file from a previous survey, click on the **I don't know. Work it out for me.** button. This will open the data file, and allow you to analyse the data, to calculate the variance and prevalence estimates needed. It will also calculate an estimate of the population size. Be sure to set this yourself if the data came from a different population. See Data Analysis on page 160 for instructions on analysing data.
- Step 8:** When all the parameters are entered, click the **Calculate** button.
- Step 9:** The results will be displayed in a window. The best first- (herd) and second- (village) stage sample sizes will be shown. The second-stage sample size will be expressed as either a number (Design 1, PPS), or a percentage of the herd population (Designs 2 and 3, SRS and RGCS).



## Step 5: Stratification

Stratification almost always improves the accuracy of the survey, and usually makes the field work simpler (see page 45). When little information is available about the population, stratification is usually done by geographical area. For instance, a national survey may be stratified by state or province, or a provincial survey may be stratified by district.

Proportional allocation

In order to make sure that each area is properly represented, we usually want to select villages from each stratum proportional to the total number of villages in that stratum. This means that a district with more villages will contribute more villages to the sample than a district with fewer villages. This is known as *proportional allocation*. The number of villages to be selected from one stratum ( $n_k$ ) is equal to the total number of villages to be selected ( $n$ ) times the proportion of villages in the population ( $N$ ) that are in that stratum ( $N_k$ ).

$$n_k = n \times \frac{N_k}{N}$$

**Example:** In a survey, the first-stage sample size is 40, and the total number of villages in the study area is 480. There are 5 districts which are used for stratification. The number of villages in the first district is 120. The proportion of the total villages in that district is  $120/480 = 1/4$ . The number of villages to be selected from that district is therefore  $40 \times 1/4 = 10$  villages. A district with 80 villages would contribute  $40 \times (80/480) = 6.67 \approx 7$  villages.

When using stratification, you can generally use the same sample size that you would calculate without stratification, and divide it up between the strata with proportional allocation. The overall results will usually be slightly more precise than predicted, due to the stratification. Note that the estimates for the individual strata will be much less precise than the overall estimate, because the sample size in each stratum is much smaller than the overall sample size. If you require precise estimates for each stratum, calculate the sample size required for each stratum separately. You can then combine the stratum results to give an overall estimate (which will be very precise because of the large sample size).

## Step 8: First-stage sampling - the herd or village

The approach used for first stage of sampling depends on the survey design used. In all cases, however, sampling is done with replacement. This means that the same herd or village can be chosen twice. In this case, twice as many animals as normal are sampled from the village.

**Example:** The calculated sample size for a two-stage prevalence survey is 40 villages and 8% animals in each village. A good sampling frame is available, but no livestock population figures, so Design 2 is used. The villages are selected from the sampling frame using simple random sampling with replacement. One village is selected twice. The sample size is still 40, even though only 39 separate villages are visited. The village that was selected twice has a population of 145 pigs. Instead of the normal 8%, two samples (16%) are drawn from this population giving a total of 24 animals.



If a sampling frame is available on computer disk, and **Designs 1 or 2** are used, you can use the **Random Village** program to select the herds or villages (see page 50). Use the following steps to select the villages:

- Start the Random Village program.
- Click on the **Open** button and select the data file containing the sampling frame.
- Click on one or more Identification Fields to be displayed for the selected villages or herds (usually ID, name etc.).
- Under Number to Select, enter the total number of villages or herds (the first stage sample size).
- Enter the sampling type. If using Design 1 with probability proportional to size (PPS) sampling, click on Probability Proportional to Size sampling. You will then need to select the field from the table that has the size information (livestock population). If you are using Design 2, select Simple Random Sampling.
- Under Replacement, click With Replacement.
- If using stratification, click the check box next to Use Stratification?. You will then need to select the field that contains the information used for stratification. This will usually be a province or district code.
- Click on the **Select** button to select the random sample.
- The selected villages or herds are displayed. You can save them to a new table, or print them.

If the sampling frame is not available on disk, you can use the manual techniques described in Chapter 3 to do either PPS (page 48) or SRS (page 41) sampling.



If no sampling frame is available, and you are using **Design 3**, you can use the random geographic coordinate sampling program (**RGCS Win95**) to select random coordinates (see page 68). If you have access to a copy of the ArcView GIS program and a digital map of the study area boundaries, you can use **RGCS ArcView** to select coordinates (see page 69).

If using random geographic coordinate sampling, you will then need to screen the selected points (if possible), and then visit points to identify nearby villages. This process is described in detail in Chapter 3 (page 74).

## Step 11: Second-stage sampling - the animal

Once herds or villages have been selected, and the field work commenced, the second stage of sampling, selecting individual animals, can be done.

In all cases, the animals are selected without replacement, so that an individual animal can be tested only once. If Design 1 (PPS) is used, then a fixed number of animals is selected from each herd or village sampled. If Designs 2 or 3 are used, a fixed proportion of the total village or herd population is selected.

If the animals are in a single herd, then a sampling frame may already exist, maintained by the owner. This can be used for simple random sampling done either by hand, or with a computer after first entering the data into the **Random Village** program. Alternatively, randomised systematic sampling (page 44) can be used if the facilities exist for putting all the animals in a sequence (e.g. handling yards and races).

In a village with multiple livestock owners, it is usually necessary to build a sampling frame first, and then select the random sample. Village interviews of livestock owners are an efficient way to build an accurate sampling frame and are described in detail in Chapter 5. The technique for selecting animals from this sampling frame using a manual method is described on page 55, and the use of the **Random Animal** program is described on page 58.

## Step 19: Data analysis

The analysis of prevalence data collected in a two-stage survey using any of the three designs is very complex. The formulae used are listed in Appendix B. The **Prevalence Analysis** program, which also calculates sample sizes for two-stage prevalence surveys, is included on the CD to analyse data from the three types of survey designs described.

Before analysis is done, the data must be entered into a computer, and stored in a file in either dBASE or Paradox format. Epi Info or another database program can be used to enter the data. The Prevalence Analysis program may also be used for simple data entry.

### Data inputs

The data files, data fields, and other information required for analysis depend on the survey design used and whether or not stratification was used. In all cases, an animal-level data file is required, with the disease status of each animal, and the village that the animal came from. The disease status may be a code which indicates Diseased/Non-Diseased, or a Yes/No field. It may also be a numeric value, such as an antibody titre. In this case you need to specify a cut-off value. Above this cut-off value, animals are considered to be positive, and below the value they are negative.



## Design 1 (Probability Proportional to Size)

### Without stratification

File 1 (Animal File)

- Disease status
- Village ID

### With stratification

File 1 (Animal File)

File 2 (Village File)

- |                  |              |
|------------------|--------------|
| • Disease status | • Village ID |
| • Village ID     | • Stratum ID |

## Design 2 (Simple Random Sampling)

### Without stratification

File 1 (Animal File)

File 2 (Village File)

Other figures

- |                  |                                |                                |
|------------------|--------------------------------|--------------------------------|
| • Disease status | • Village ID                   | • Total villages in study area |
| • Village ID     | • Village livestock population | • Total animals in study area  |

### With stratification

File 1 (Animal File)

File 2 (Village File)

File 3 (Stratum File)

- |                  |                                |   |
|------------------|--------------------------------|---|
| • Disease status | • Village ID                   | • Total number of villages in the stratum |
| • Village ID     | • Village livestock population | • Total number of animals in the stratum  |
|                  | • Stratum ID                   | • Stratum ID                              |

## Design 3 (Random Geographic Coordinate Sampling)

### Without stratification

File 1 (Animal File)	File 2 (Village File)	Other figures
<ul style="list-style-type: none"> <li>• Disease status</li> <li>• Village ID</li> </ul>	<ul style="list-style-type: none"> <li>• Village ID</li> <li>• Village livestock population</li> <li>• Village weight<sup>9</sup></li> <li>• Area fraction<sup>10</sup> (optional)</li> </ul>	<ul style="list-style-type: none"> <li>• Selection radius</li> <li>• Total number of random points used</li> <li>• Total area of study area</li> </ul>

### With stratification

File 1 (Animal File)	File 2 (Village File)	File 3 (Stratum File)
<ul style="list-style-type: none"> <li>• Disease status</li> <li>• Village ID</li> </ul>	<ul style="list-style-type: none"> <li>• Village ID</li> <li>• Village livestock population</li> <li>• Village weight</li> <li>• Stratum ID</li> <li>• Area fraction (optional)</li> </ul>	<ul style="list-style-type: none"> <li>• Stratum ID</li> <li>• Selection radius for each stratum</li> <li>• Total number random points used in the stratum</li> <li>• Total area of the stratum</li> </ul>

Some examples of appropriate Questionnaire file formats for Epi Info are shown below.

Demonstration Data Entry Form Prevalence Survey - Animal Data		
Tube Number:	#####	
Village ID:	#####	
District ID:	#####	(If using district for stratification)
Age:	##	
Sex:	<A>	(Single character, uppercase field)
Species:	_____	(E.g. cattle / buffalo, if necessary)
Antibody titre:	#####	(As reported by laboratory)
Disease status:	<Y>	(Calculated from antibody titre) (Use a standard cut-off value)

<sup>9</sup>The *village weight* is the total number of villages within the selection radius of the point used to select the village.

<sup>10</sup>The *area fraction* is the proportion of the area of the circle (as defined by the sampling radius around the random point used to select that village) which lies inside the study area. For most villages, this will equal 1, but for some near the boundary of the study area, it will be smaller.

Demonstration Data Entry Form  
Prevalence Survey - Village Data  
(For Design 2 (SRS) and 3 (RGCS) only)  
(Not required for design 1 (PPS))

Village ID: #####

Total population: #####

District ID: ##### (If using district for stratification)

Demonstration Data Entry Form  
Prevalence Survey - Village Data  
(For Design 3 (RGCS) only)  
(Not required for design 1 (PPS) and 2 (SRS))

Village ID: #####

Total population: #####

District ID: ##### (If using district for stratification)

Weight: ## (Total number of villages around the point)

Selection Radius: ##.## km (must be the same for every village in one stratum)

Total points: ## (total number of points used, including points with no villages. Same for every village in one stratum)

Study area: #####.## sq km (Total area or stratum area. Same for every village in one stratum)

## Analysing the data

To analyse the data use the following steps:

- Step 1:** Start the **Prevalence Analysis** program by clicking on the Windows **Start** button, selecting Programs, then Survey Toolbox and choosing Prevalence Analysis.
- Step 2:** Click on the Prevalence Data Analysis tab at the top of the window. The other tab is for sample size calculation.
- Step 3:** In the First Stage Sampling Scheme box select the survey design used, Probability Proportional to Size (Design 1), Simple Random Sampling (Design 2) or Random Geographic Coordinate Sampling (Design 3).
- Step 4:** If stratification was used, click the Stratification? check box.
- Step 5:** In the data fields, open the files and enter the data required for the type of analysis you are performing.
- Step 6:** In the Animal data box, Click the **Open Animal Data** button, and select the file with the animal level survey data.

**Sample size and data analysis** X

**Two-Stage Prevalence Surveys** *Survey Toolbox*

Sample Size Calculation | Prevalence Data Analysis

First Stage Sampling Scheme

Probability Proportional to Size Sampling (PPS)

Simple Random Sampling (SRS)

Random Geographic Coordinate Sampling (RGCS)

Stratified?

Second Stage Sampling Scheme

Fixed proportion of population randomly selected

Data Fields

Animal Data

Open Animal Data | Edit

Disease Status: AB\_POS

Village ID: VILLAGE

Status Codes

Positive: True

Negative: False

Village Data

Open Village Data | Edit

Village ID: VILLAGEID

Population: CATTLE

Stratum ID: STRATUM

Stratification

Open Strata Data | Edit

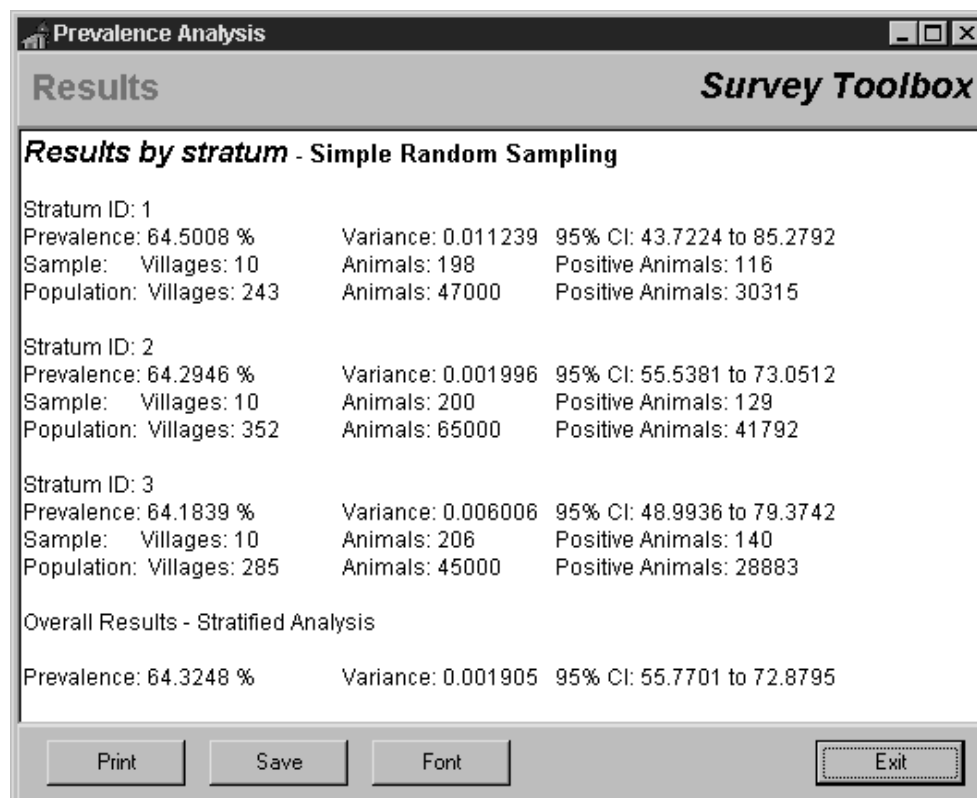
Stratum ID: STRATUM

Total Villages: VILLAGES

Total Animals: CATTLE

Calculate | Exit

- Step 7:** In the data fields box, select the fields in the database that contain the data for analysis. First, select the field that contains the disease status data.
- Step 8:** Then select the field which contains data identifying which village or herd the animal came from (First stage sampling units).
- Step 9:** Make sure the codes for disease status in the Status Codes box are correct.
- Step 10:** If displayed, enter data in the Village Data box. Click on the **Open Village Data** button, and select the fields required.
- Step 11:** If further information is required for stratification or random geographic coordinate sampling, enter the data required. Click on the **Open Strata Data** button, and set up the fields, or type in the parameters required.
- Step 12:** When all the fields have been entered, click on the **Calculate** button to analyse the data. A window will display the results, which can be either printed or saved to a file.



## Calculating true prevalence

The disease status of animals during a survey is assessed by means of a laboratory test, or by direct clinical examination. In both situations, it is possible to make a mistake in a few cases, and call some healthy animals diseased, and some diseased animals healthy (or alternatively, think that some animals with antibodies don't have them, and others that don't, do have them). There are two measures used to describe how good a test is at correctly determining the disease state of an animal: sensitivity and specificity (see page 33 for a full discussion).

Because most tests make a few mistakes, a few of the test results analysed could be wrong, which makes the estimate of the prevalence incorrect. Usually this error is quite small, but for tests that make mistakes more often, the error can be large.

If the sensitivity and specificity of the test are known or can be estimated, it is possible to correct for these mistakes, and convert the results of the analysis, the *apparent prevalence*, to the corrected result, the *true prevalence*.

The **True Prevalence** program on the CD carries out the calculations for you. When the results have been analysed with the Prevalence Analysis program, use True Prevalence to convert the result to the true prevalence, based on the test sensitivity and specificity:



- Step 1:** Start the True Prevalence program. Use the Windows Start menu, select Programs, Survey Toolbox, True Prevalence.
- Step 2:** In the Parameters box, enter the Apparent Prevalence, as reported by the Prevalence Analysis program.

- Step 3:** Enter the test sensitivity and specificity. The laboratory may be able to suggest figures for these, or else it might be necessary to search journals for published studies.
- Step 4:** Enter the sample size of the survey.
- Step 5:** Press the **Calculate** button.
- Step 6:** The true prevalence is shown, along with a confidence interval.

**Note:** The confidence interval is based on the assumption that the sample was selected by single-stage simple random sampling. For the two-stage surveys described in this chapter, the confidence interval reported will be smaller than the correct confidence interval.

Parameters	
Apparent Prevalence	14 %
Sensitivity	94 %
Specificity	88 %
Sample Size	560

Result	
True Prevalence	2.439 %
95% Confidence Interval	(0.000 - 5.944)

## Interpretation of results

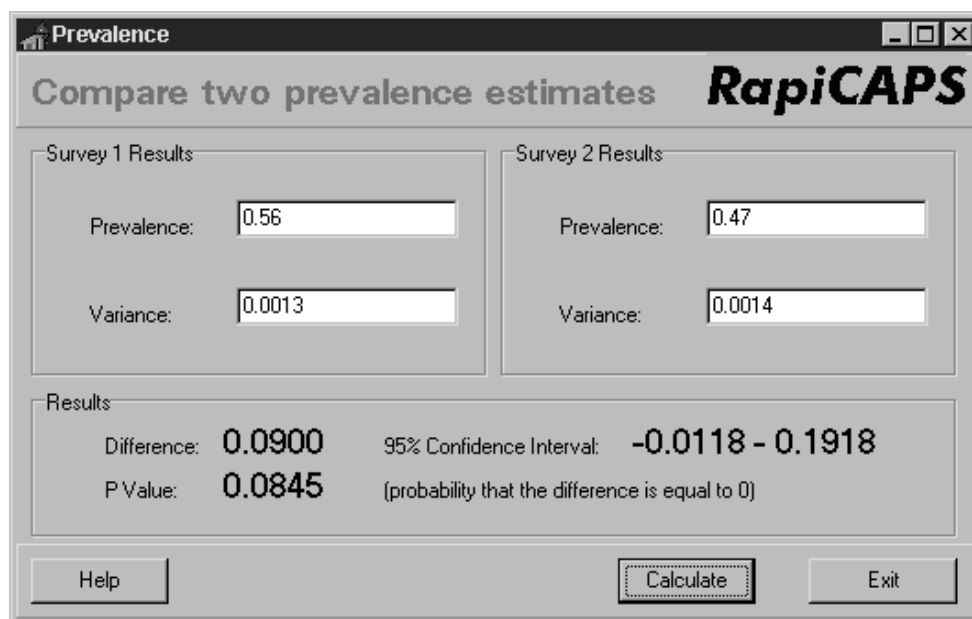
The key result from the survey is an estimate of the prevalence of the disease or state in the population. This is shown as a single figure, and a 95% confidence interval. The confidence interval can be interpreted to mean: "If the same survey were conducted in the same population many times, the confidence interval produced by the results would include the true prevalence of disease in the population 95% of the time." This can be loosely interpreted to mean that we are 95% confident that the true prevalence lies within the confidence interval.

If stratification was used, separate prevalence estimates and confidence intervals are shown for each of the strata. Note that because the number of animals sampled from each stratum is relatively small, the confidence intervals for the stratum estimates are usually very wide, indicating that our estimates are not very precise. The overall estimate is usually much more precise with narrow confidence limits.

### Comparison of two prevalences

When the prevalence estimates from two surveys have been calculated they can be compared to determine whether there is a real difference between them, or the difference is likely to be just due to chance. Monitoring changes in prevalence is an important way to evaluate the progress of a disease control program.

Use the **Compare Prevalence** program to compare two prevalence estimates. Click on the Windows Start menu, then select Programs, Survey Toolbox, Compare Prevalence.



Survey	Prevalence	Variance
Survey 1	0.56	0.0013
Survey 2	0.47	0.0014

Results	Value
Difference	0.0900
95% Confidence Interval	-0.0118 - 0.1918
P Value	0.0845

- Step 1:** In the Survey 1 Results box, enter the prevalence and variance from the first survey, as reported by the Prevalence Analysis program.
- Step 2:** Enter the same figures from the second survey in the Survey 2 Results box.
- Step 3:** Click on the **Calculate** button.
- Step 4:** The results will be displayed.

The results show the difference between the two prevalence estimates, and a 95% confidence interval for that difference. In addition, they show a P value, which is a measure of the probability that the two prevalence estimates are in fact the same (the difference is 0). If the P value is very small, then we can be confident that there is a real difference between the two prevalences.

# 8

## Incidence Rate Surveys

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Disease outbreak surveys

Introduction

Survey procedure

Sample size

Sampling (selecting herds or villages)

Data collection - the interview

Data management

Data analysis

Interpretation of results

Analysis of two data sources

Background

Data sources

Sampling

Data collection

Data management

Data analysis

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



## Incidence rate

Incidence rate is the number of the new cases of disease in a population at risk over a period (see page 28). Incidence rate measures the rate of spread of infectious diseases. It can also help distinguish between naturally acquired antibodies and vaccine-induced antibodies when used in conjunction with a prevalence survey.

One way to estimate incidence rate is to observe a group of animals for a long period of time, and to record which animals become affected with the disease. This type of incidence rate study is very slow and expensive, as every animal has to be regularly examined, and the study may last for many months or longer. These kinds of study have commonly been carried out in developed countries.

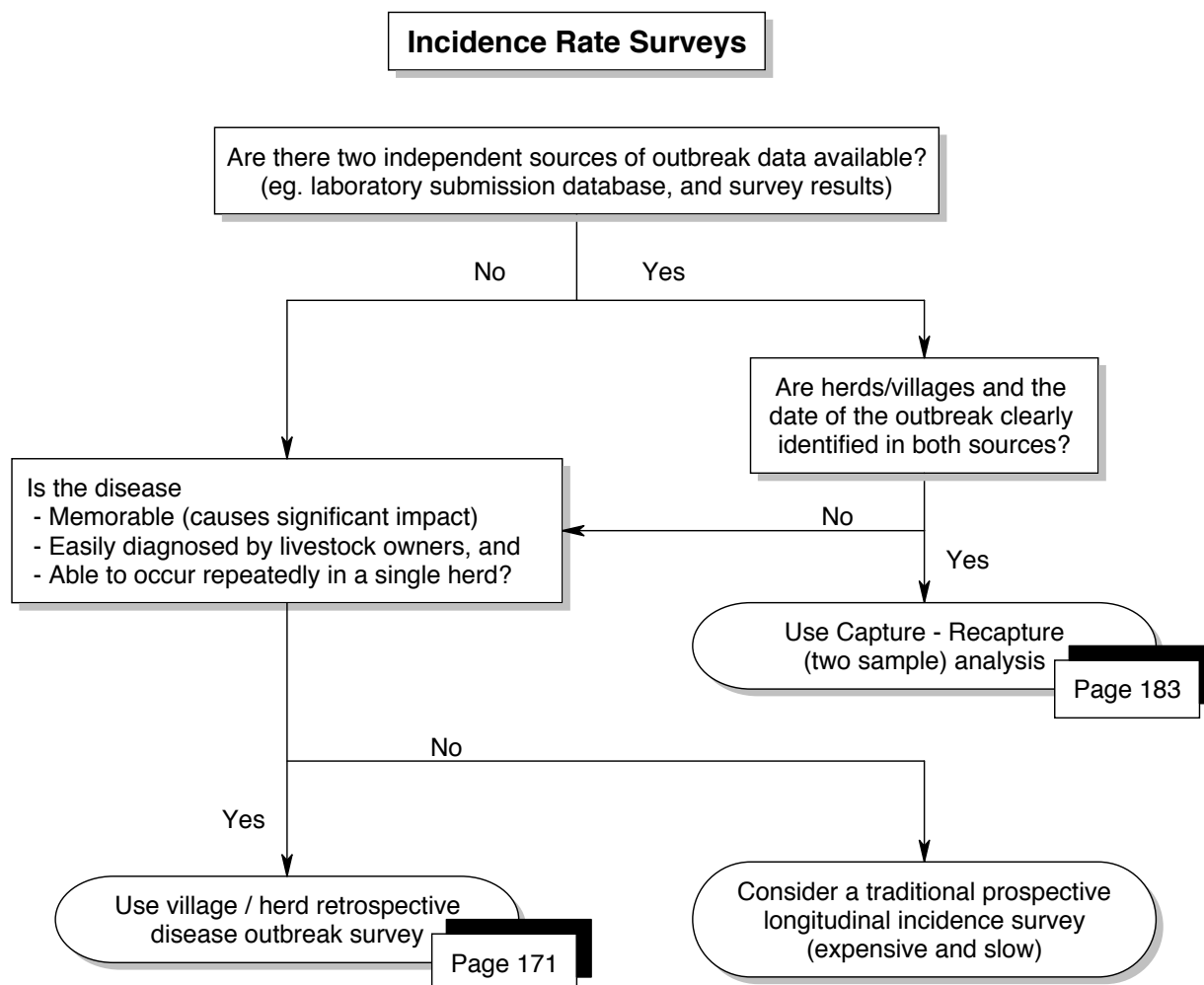
This chapter describes two alternative ways to gather information about disease incidence rate. The differences between these techniques and traditional studies are:

- the unit of interest is the village or herd, instead of the animal. This means that we are not interested in the number of animals that get a disease over a period of time, but the number of villages or herds that suffer an outbreak of disease over a period of time;
- the outbreak information is collected by using the memories of the livestock owners. Instead of starting a study, and observing the animals for a long time (a *prospective* study), livestock owners are asked about the disease outbreaks that have occurred over several years before the survey (a *retrospective* study).

For most major epidemic diseases, the disease is maintained through spread from one village or herd to another. It is rare that the bacteria or viruses are able to maintain themselves within the one herd, as animals either die, or develop immunity relatively quickly. Herd- or village-level incidence rate is therefore a more useful measure of the rate of spread of disease, or the effectiveness of a control program, than individual animal-level incidence rate. In a survey depending on the memories of livestock owners, it is also much easier to collect reliable information about herd- or village-level outbreaks than about disease in individual animals.

The first of the two techniques described in this chapter is the Retrospective Disease Outbreak Surveys technique. Village interviews are carried out to ask livestock owners about previous disease outbreaks. When the results are analysed, the measure of disease is not a traditional incidence rate measure, but can be used in the same way to assess the rate of spread of disease, and compare disease levels between different areas (or the same area at different times).

The second technique, Capture-Recapture technique, uses information on disease outbreaks that the veterinary authorities may have already collected. This technique uses two sources of outbreak data (such as diagnostic laboratory submissions or field disease reports), and combines them to estimate a traditional incidence rate measure. If two separate suitable data sources already exist, no field survey is necessary.



## Retrospective disease outbreak surveys

### Introduction

In this technique interviews with herd or village livestock owners are carried out to collect information about the date of the most recent disease outbreak. The reliability of the survey therefore depends on the livestock owners' ability to correctly identify the disease, and to correctly recall the date that the outbreak started.

In order to ensure that the quality of information collected is high, the survey technique should be used only in appropriate situations. Firstly, it should be used to investigate only diseases which are:

#### Survey prerequisites

- *discrete and repeatable*: The disease must occur in outbreak form, last for a relative short time and it must be possible for it to occur more than once in the same village.

- *distinctive and well known*: Diagnosis is based entirely on the observations of the livestock owners. Diseases that are clearly different to other diseases, and have a dramatic and consistent clinical presentation are more easily diagnosed.
- *memorable*: The ability of livestock owners to remember the date of an outbreak depends on the effect the disease had on them. The more dramatic the disease, and the more disruption to the lives of the livestock owners, the more reliably it will be remembered.

Every effort is made to assist the livestock owners to accurately remember the date of the outbreak. A range of techniques available is discussed in Chapter 5.

The various strengths and weaknesses of the technique are summarised below.

Strengths and weaknesses of the Retrospective Disease Outbreak Survey methodology for estimating disease occurrence.

Strengths	Weaknesses
Rapid - collects data retrospectively	Data accuracy - depends on recall
Group interviews may be used to collect other data simultaneously	Limited to diseases causing significant impact and occurring in cyclic epidemics
Can be used for quantitative comparisons	Does not provide direct estimate of incidence rate
Inexpensive - no laboratory test or repeat visits	Requires staff training
	Depends of owner diagnosis
	No animal-level estimates

## Survey activities

The major steps in carrying out a Retrospective Disease Outbreak Survey are:

- Step 1:** Identify the question to be answered (disease and geographic area of interest).
- Step 2:** Identify the target population.
- Step 3:** Decide if the survey is to use stratification.
- Step 4:** Calculate the best sample size.
- Step 5:** Plan field activities.
- Step 6:** Train survey teams.
- Step 7:** Pilot survey.
- Step 8:** Select the sample.
- Step 9:** Visit selected herds or villages.
- Step 10:** Hold a livestock owner interview.
- Step 11:** Determine if the herd or village has ever had an outbreak of the disease.
- Step 12:** If so, determine the date of the start of the last outbreak.

- Step 13:** If not, determine the earliest date since which the livestock owners are confident that there has been no outbreak.
- Step 14:** Check the data for completeness and accuracy.
- Step 15:** Enter the survey data into a computer.
- Step 16:** Check the data for mistakes during data entry.
- Step 17:** Recode the data ready for analysis.
- Step 18:** Analyse the data.
- Step 19:** Report the data.

The key steps are described in detail below.

## Step 4: Sample size

Unlike the prevalence surveys described in Chapter 7, the unit of interest in Retrospective Disease Outbreak Surveys is not the individual animal, but the herd or village. This means that simple one-stage sampling can be used.

The sample size is calculated on the basis that the survey will be used to compare the rate of disease outbreaks, either in two different areas, or perhaps more importantly, in the same area at two different times. For example, if a Retrospective Disease Outbreak Survey was conducted last year, and then repeated this year, after the introduction of a disease control program, comparing the level of disease may indicate the success of the program.

The measure that is used to calculate sample size is median (or mean) time since the last disease outbreak. If all villages or herds in the survey have experienced the disease this is simply the average of the times since the last outbreak. When comparing the results of two surveys, if the median time since the last outbreak is large, only a relative small number of herds or villages are needed to be sure that this difference is not due to chance. If the difference in median times is very small, and the two groups are almost the same, many more herds or villages are needed to determine if the small difference is real, or just due to chance.

Use the **Survive Size** program to calculate the sample size needed for the survey. To start the program, use the Windows Start menu, select Programs, then Survey Toolbox, and Survive Size. Calculate the sample size using the following steps:



- Step 1:** In the box labelled Estimated Median Survival Times, enter the times for group 1 and for group 2. You can type in the times in any units (months, years, or days). These times are what you expect to see from the survey. Alternatively, you can think of these times as indicating the smallest real difference you want to be able to detect. If the difference is smaller, then your survey will not be able to reliably distinguish between them.

**Example:** An FMD vaccination program has been started in one part of a country where the disease is endemic. In order to monitor the progress of the vaccination campaign, it is decided to carry out disease outbreak surveys every year. Before the program started, the average time since the last outbreak for villages in the area was about 3 years. The veterinary authorities decide that lengthening this to 5 years is a good indication that the program is being successful, but anything less could be accounted for by year to year variation. When calculating the sample size for the survey,

they use 3 and 5 years as the median survival times for group 1 and group 2.

- Step 2:** In the Parameters box, enter the Significance you want. This is indicating how confident you want to be in the result<sup>11</sup>. Usually, you can leave this as 95%
- Step 3:** In the Parameters box, enter the Power you want. This is a measure of how well the survey will be able to determine if there is a difference between the two groups<sup>12</sup>.
- Step 4:** Click the **Calculate** button and the sample size will be displayed.

The screenshot shows a software window titled "Survival Analysis" with a sub-window "Sample Size Calculation" from the "Survey Toolbox". It contains input fields for "Estimated Median Survival Times" (Group 1: 3.5, Group 2: 5.5) and "Parameters" (Significance: 95%, Power: 90%). The "Results" section displays "Number of events (outbreaks) required in each group: 103". Buttons for "Help", "Calculate", and "Exit" are at the bottom.

The sample size is the number of herds or villages that have had an outbreak that need to be included in each group. During the survey, some villages may not have had an outbreak, so the overall sample size will need to be slightly larger than the number suggested by the program. When deciding on the sample size, you can either use your experience or judgment to estimate what proportion of herds or villages have never experienced an outbreak. Another approach is to carry out the survey, and continue to select more herds or villages until enough have been visited.

Most importantly, as with all sample size estimates, the figures should only be used as a guide, or rough estimate.

## Step 8: Selecting herds or villages

There are two ways to select the herds or villages: 1) simple random sampling (SRS), and 2) random geographic coordinate sampling (RGCS). Firstly, if a sampling frame (page 49) is available, the easiest approach is to use simple random sampling from that sampling frame. To select villages manually, use the procedure described on page 41.

<sup>11</sup>Significance is the probability that the results of the survey will indicate that there is no difference between the two groups when the two groups are the same.

<sup>12</sup>Power is the probability that the results of the survey will indicate that there is a difference between the two groups when there actually is.



If the sampling frame is available on computer, you can use the **Random Village** program to select the sample, as described on page 50, using the following settings:

- Sampling Type should be set to Simple Random Sampling (probability proportional to size sampling is not appropriate here).
- Replacement should be set to Without Replacement.
- Do not select stratification.

If no sampling frame is available, you can select herds or villages using random geographic coordinate sampling (RGCS). This technique is described in detail in Chapter 3. When using RGCS, be sure to record village or herd weights, for use during analysis.

## Steps 10 - 13: The interview

Village livestock owner interviews are discussed in detail in Chapter 5, including advice on techniques for collecting information about the date of the last disease outbreak (page 118). It is worth repeating the importance of establishing a censoring time for villages that have not experienced an outbreak (page 120).

In addition to these two dates, it is possible to collect other related information to help with the analysis. This may include:

- the number of animals in the village at the time of the interview;
- the number of animals in the village at the time of the outbreak (or the censoring time); and
- the proportion of those animals that were affected by the outbreak.

An example data recording sheet is shown in Appendix D.

## Step 15: Data management

When the field work is completed all the results need to be entered into a computer for analysis. See Chapter 6 for general advice on computerised data management.

Data may be entered using any database program that can export data to dBASE or Paradox format (including Epi Info). When creating the database table, the following fields are necessary:

- Village or herd identification
- Outbreak (yes/no, or code field indicating whether or not there has ever been an outbreak).
- Date of last outbreak (or censoring time). This may be included as a single date field if the day of the start of the outbreak has been estimated. Usually, it is only possible to recall the month of the outbreak. There are two solutions to this problem. Firstly, all outbreaks can be arbitrarily said to have started on 15<sup>th</sup> of the month. Alternatively, two numeric fields could be used, one for the month, and one for the year.
- Date of visit. This could be treated in the same way, with either a date field, or only month and year recorded in two separate fields.
- Time since outbreak. This field is left blank at data entry, and is calculated from the two dates later.

- If random geographic coordinate sampling was used to select the villages, there must also be a field to record the weight for each village.
- If the data from two areas or times are being compared, then the file needs to contain all the data from the two groups. If the data already exist in two separate tables, they can be merged into a single table using the merge procedure (available in most database programs, including Epi Info – see the Epi Info on-line manual). There must be a Group field, with a code to indicate which group the record belongs to. This may be a numeric code, text field, or a yes/no field.

An example questionnaire file for creating the table in Epi Info is shown below.

```

                                Demonstration Data Entry Form
                                Village Outbreak Survey

Village ID:      #####

Date of visit:  Month ##   Year ####

Had Outbreak?  <Y>           (Censoring variable)

Outbreak:      Month ##   Year ####
(or censoring time)

Time since outbreak:  ##.### (Calculated from visit date
                                and outbreak date)

Weight:        ## (Random geographic coordinate sampling only)
                                (Number of villages within selection radius)

Group          #           (when comparing two groups only)
    
```

Other fields can be included for more complex analysis, such as livestock population at the time of the visit, population at the time of the outbreak, and proportion of animals affected. Analysis of these extra data requires more sophisticated techniques, and possibly specialised statistical programs which are not provided with the Survey Toolbox software. Complex analysis and software (described later) are not necessary to calculate the level of disease and compare two groups.

**Epi Info**

Before analysis, the dates recorded must be used to calculate the time since the last outbreak. The exact procedure depends on the database program being used, but most are similar. When using Epi Info, the procedure is as follows:

- Step 1:** Start Epi Info, and choose Analysis from the programs menu.
- Step 2:** Open the data file, using the Read command. If the file is in dBASE format, use Read \*.dbf. Select the file from the list.
- Step 3:** If the outbreak time and visit time have been stored as Date fields, calculate the time since last outbreak using the command: Time = Date1 - Date2

**Example:** If the date of the outbreak is stored in a field called OBDate, the date of the visit is VisDate, and the time since the last outbreak is to be stored in a field called Time, then type:  $\text{Time} = \text{VisDate} - \text{OBDate}$ . The result is the number of days between the two dates. If it does not already exist, you will have to create a new Time field, using the **define** command:  
**define time ###.**

**Step 4:** If the outbreak and visit times have been stored in separate month and year fields (for instance, OByear and OBmonth for the outbreak time, and Vyear and Vmonth for the visit time) use the command:  $\text{Time} = (\text{Vyear} - \text{OByear}) + ((\text{Vmonth} - \text{OBmonth}) / 12)$ . This will give the outbreak time in terms of years. If months are preferred multiply by 12.

**Step 5:** Save the new values to a new data file using the **Route** and **Write recfile** commands.

**Example:** To save the data in a file called obsurvey.rec, use the commands **Route obsurvey.rec** and then **write recfile**

## Step 18: Data analysis

### Basic analysis

The data are analysed using a special technique called *survival analysis*. This technique uses the times that a herd or village *survives* without experiencing an outbreak. The advantage of survival analysis is that it is possible to include data from herds or villages that have not had an outbreak.

The Survival program, included in the Survey Toolbox does all the analysis necessary. To analyse the data using the program:



**Step 1:** Start the Survive program. Use the Windows start menu, select Programs, Survey Toolbox, and choose Survive.

**Step 2:** Open the data file for analysis. Click the **Open** button, and select the file from the list. The file must be in dBASE or Paradox format. You can also use the program to create a new file containing all the data you need, by clicking the **New** button.

**Step 3:** You now need to tell the program which fields the data are in. In the Data Fields box, click on the arrow at the right of the Survival Times box. Select the field that stores the time since the last outbreak.

**Step 4:** Next click on the Censoring Indicator box, and select the field that indicates if the herd or village has ever had an outbreak.

**Step 5:** You then have to tell the program what the codes in the field mean. The computer displays the codes from the censoring field in the Censoring Codes box. Censored is the code for villages that have never had an outbreak. Uncensored is the code for villages that have had an outbreak. If the codes are the wrong way around, click on the **Switch** button to swap the codes.

**Step 6:** If the herds or villages were selected using random geographic coordinate sampling, click on the Weighted? checkbox, and select the

field that contains the village or herd weights. If you selected villages using simple random sampling, you can leave this unchecked.

- Step 7:** Select the type of analysis that you are performing. If you are analysing the results of a single survey, and not doing any comparisons, select Single Group Analysis. If data from two groups are in the data file, you can select just one group by clicking on the Select Group check box.
- Step 8:** If you are comparing two groups, select Compare Two Groups. You then need to say which field the group identifier is stored in. Click the Grouping Variable box, and select the field. The program will check the codes in the file, and display them in the Group Codes box. You can use the **Switch** button to swap the codes from group 1 to group 2.
- Step 9:** If comparing two groups, you may want to adjust for seasonal differences (see below). Click the check box to adjust for differences.
- Step 10:** When all the fields have been set, click on the **Analyse** button, and the results will be displayed. You can print or save the results of the analysis.

### Adjusting for seasonal differences

Some diseases have a clear seasonal pattern, with more disease at one time of year than another. If this is the case, the time of year that the survey is conducted will have an effect on the length of time since the last outbreak. If the survey is conducted just after the peak season for disease outbreaks, many herds or villages may have experienced an outbreak in the last two or three months. If the survey is conducted 7 or 8 months after the peak season, the time since the last outbreak is longer. However, this is not because the disease situation is different. It is just because the outbreaks occur in a clear season.

If there is evidence that the disease outbreaks do have a seasonal pattern, you may have to correct for this, so the analysis is not misleading. You only need to do this if the two surveys being compared were conducted at different times of the year. If they were conducted at the same time, or during the same months in different years, no adjustment is necessary.

To adjust use the following procedure:

- Click on the Adjust for Season check box.
- In the Group Codes box, enter the month of the survey for groups 1 and 2.
- After setting up all the other fields, click the **Analyse** button to carry out the analysis.

### Complex analysis

The analysis described is usually adequate for most purposes. However, it is possible to do slightly more advanced analysis, to investigate the behaviour of the disease in more depth.

One problem when measuring herd- or village-level incidence rate is that not all herds or villages are the same size. A large village is likely to have more animals being brought in from outside the village, and this is one of the major risks for spread of disease. We could therefore expect that there would be more outbreaks in larger villages than smaller villages. When comparing two groups, if one group has more larger villages, and the other has mostly smaller villages, we would expect to see more outbreaks in the first group. This is due only to the size of the villages, not the overall level of disease. It is possible to adjust for differences in village size (livestock population), to give a fairer comparison of the amount of disease in the two groups. There are two ways to do this.

The first is use a different measure of time. Normally, we say that a village has been at risk of having an outbreak for a certain number of years. Instead, we could consider that each animal in the village has been at risk of getting disease for that period. The *animal-time* (number of animals in a village or herd, times the length of time since the last outbreak) will be greater for larger herds than small herds. Animal-time is an alternative time measurement that takes into account the size of the village livestock populations. Ideally, animal-time should be calculated on the basis of the average number of animals in the village during the time between the outbreak and the visit.

**Example:** A Retrospective Disease Outbreak Survey was carried out in 40 villages. The information collected was: 1) the time since the last outbreak (Time), 2) the pig population at the time of the visit (Vpop), and 3) the pig population at the time of the last outbreak (OBpop). In order to calculate the animal-time, the average number of animals in the village over that period was multiplied by the time since the last outbreak:  $((Vpop + Obpop)/2) \times Time$ .

Once animal-time has been calculated, the analysis can be repeated, using the animal-time field, rather than the time field. The results may show either an increased or decreased difference between the two groups, but either way, the difference will be taking variability in village livestock population into account, and therefore be somewhat more reliable.

The other approach to the analysis taking livestock population into account is to analyse the data using Cox's proportional hazards model. This is available only in sophisticated statistical software, and requires a good understanding of survival analysis and multiple regression models. If the software and technical expertise are available, village livestock population at the time of the visit and at the time of the outbreak can be included in the model.

## Interpretation of results

Unlike prevalence or traditional incidence rate estimates, the results of survival analysis are not expressed as a single number, but the estimate of a survival curve, or graph, which shows the disease experience of the entire study population. The Kaplan-Meier survival curve (named after the people that developed it) is a graph which shows the proportion of herds or villages that have 'survived' (not had an outbreak) for a particular time (measured backwards from the time of the survey). A population with fewer outbreaks or outbreaks occurring less often will have a greater proportion surviving for a longer period, so the curve will be closer to the top right of the graph. A population in which there have been recent outbreaks in most villages (from which we can imply that outbreaks are occurring frequently) will have a survival curve closer to the bottom left of the graph.

For both single and two-group analysis, the survival curves are displayed and can be printed. With experience, it is possible to interpret a survival curve. However, some summary measures which describe the curve are easier to understand.

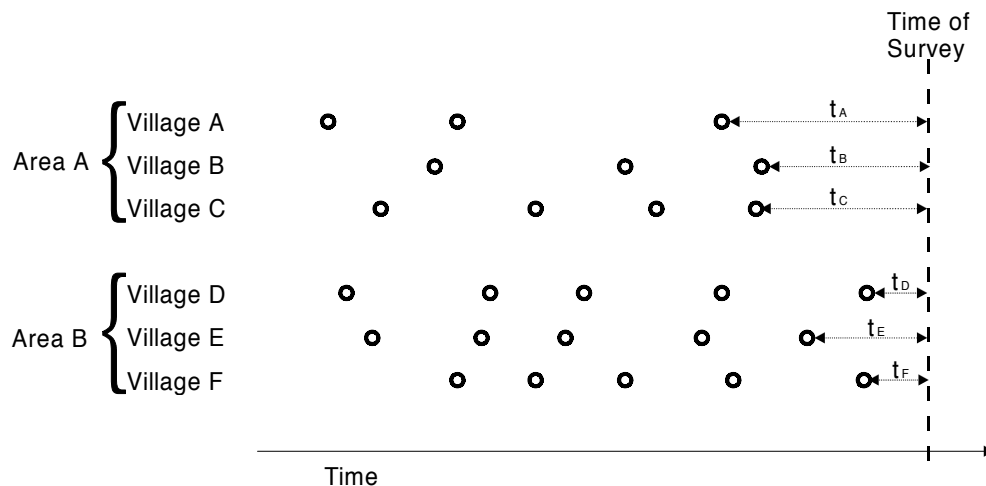
### Single group analysis

With single group analysis, the total number of observations and the total number of censored and uncensored observations are displayed. If weighted analysis is performed (because of the selection of herds or villages with random geographic coordinate sampling), these totals reflect the sum of the weights, rather than the true sum. The true totals are also displayed.

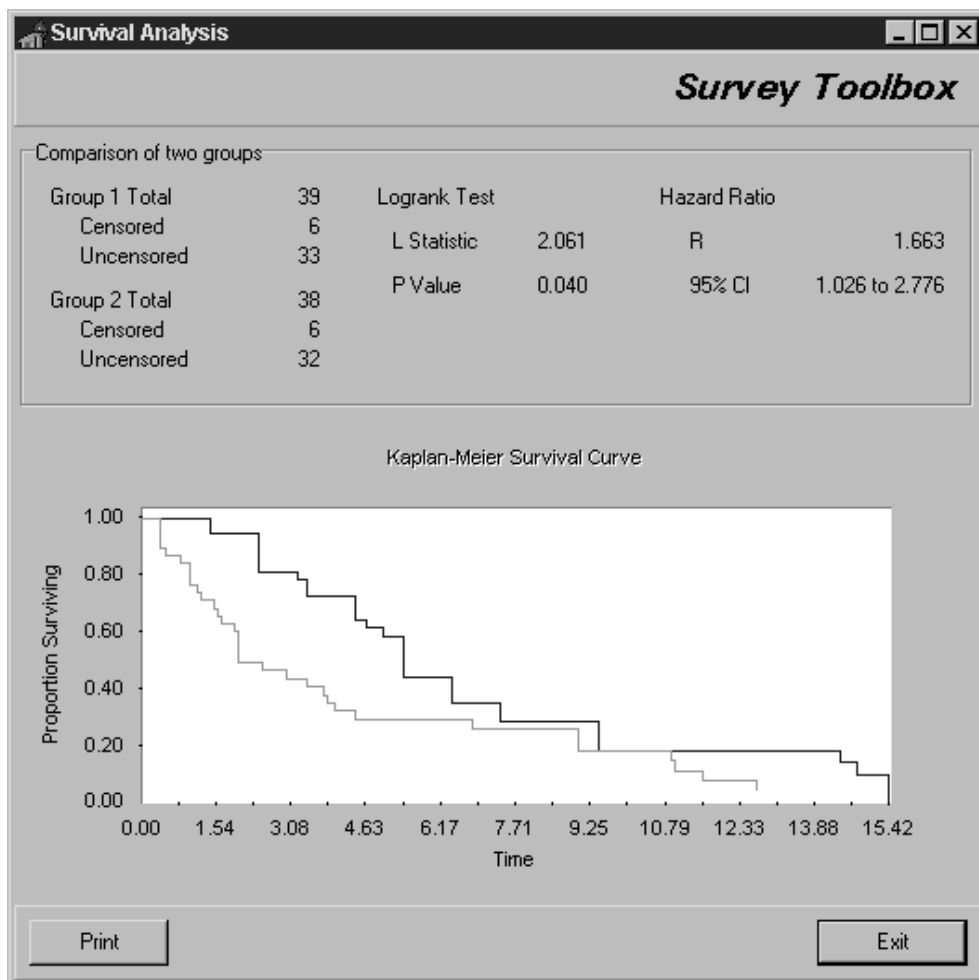
Median survival time

The *median survival time* is also shown. This is the time since which half of the herds or villages have not had an outbreak and half have. An estimate of the median survival time is used for sample size calculations (page 173).

The *mean* or *average survival time* is the average amount of time since the last outbreak for all villages. If the longest time is censored (hasn't had an outbreak, and is confident there hasn't been an outbreak for a long time), then the true mean can't be calculated. Instead, the time-limited mean is displayed, showing the mean of all times, up to a certain limit.



In the diagram, each circle represents an outbreak. The mean survival time for the villages in area A is given by  $(t_A + t_B + t_C)/3$ . The mean survival time for the villages in area B is given by  $(t_D + t_E + t_F)/3$ . The mean survival time for area A is longer than area B.



**Comparison of two groups**

The median and mean provide a basic description of the survival curve, and can be used for very simple comparisons. However, they really compare only the curve at one point. The reason for conducting a disease outbreak survey is generally to evaluate differences or changes in the disease situation. To achieve this, two groups of data must be compared.

When two groups are compared, a summary of the total number of observations in each group is presented.

Logrank test

This is followed by the *Logrank test*. This is a statistical test which compares the two survival curves to determine if there is any real difference between them, or if the apparent differences are simply due to chance.

The result of a Logrank test is given as a probability measure, called the P value. The P value is the probability that the two curves are in fact the same, or that any differences between the two groups is just due to chance. A small P value (less than 0.1 or 0.05) means that it is very unlikely that any difference is due to chance, so is interpreted as providing strong evidence that there is a real difference between the two curves. When the P value is small, there is said to be a statistically significant difference between the two curves.

The Logrank test calculates the probability that the two curves are the same. A low P value suggests that the curves are different.

Hazard Ratio

The Logrank test doesn't tell us what sort of difference there is, or how big the difference is. The key measure when comparing two groups is the *Hazard Ratio*. This is the ratio of the estimated hazard or risk of an outbreak in the two groups. A hazard ratio of 1 means that the risk of disease outbreaks in the two groups is about the same. A hazard ratio of 5 means that the risk of an outbreak in group 1 is 5 times greater than the risk in group 2. If group 1 is the population of pigs in a province two years ago, before the start of a vaccination campaign, and group 2 is the same population now, after 2 years of vaccination, a hazard ratio of 5 would mean that the risk of an outbreak amongst pigs was 5 times greater 2 years ago than it is now. This would provide strong evidence for an improvement in the disease situation. The hazard ratio is presented with a 95% confidence interval. Loosely speaking, we can be 95% sure that the true hazard ratio lies within this confidence interval.

The hazard ratio measures the ratio of the risk of an outbreak in two groups

## Analysis of two data sources

The Retrospective Disease Outbreak Surveys described above do not produce a traditional measure of incidence rate. This is because the total number of villages or herds that have suffered an outbreak in a fixed period of time can't be calculated (see page 28). This section describes a simple approach to estimating village or herd-level incidence rate which takes advantage of data that already exist.

The veterinary services in most countries maintain good records of disease outbreaks, particularly for important or notifiable diseases. This passively acquired data (compared with active surveillance – see page 14) can provide some picture of the disease situation. However, reporting is almost always incomplete, so any incidence rate estimate based on these data will be too low.

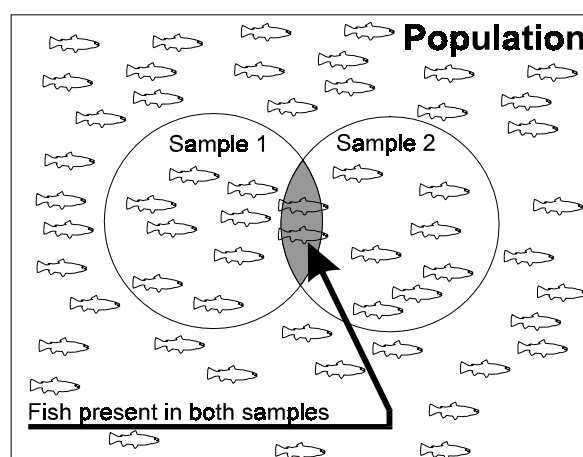
By combining this type of data with a separate, independent data source, it is possible to estimate how many outbreaks have been missed, and therefore what is the total number of disease outbreaks.

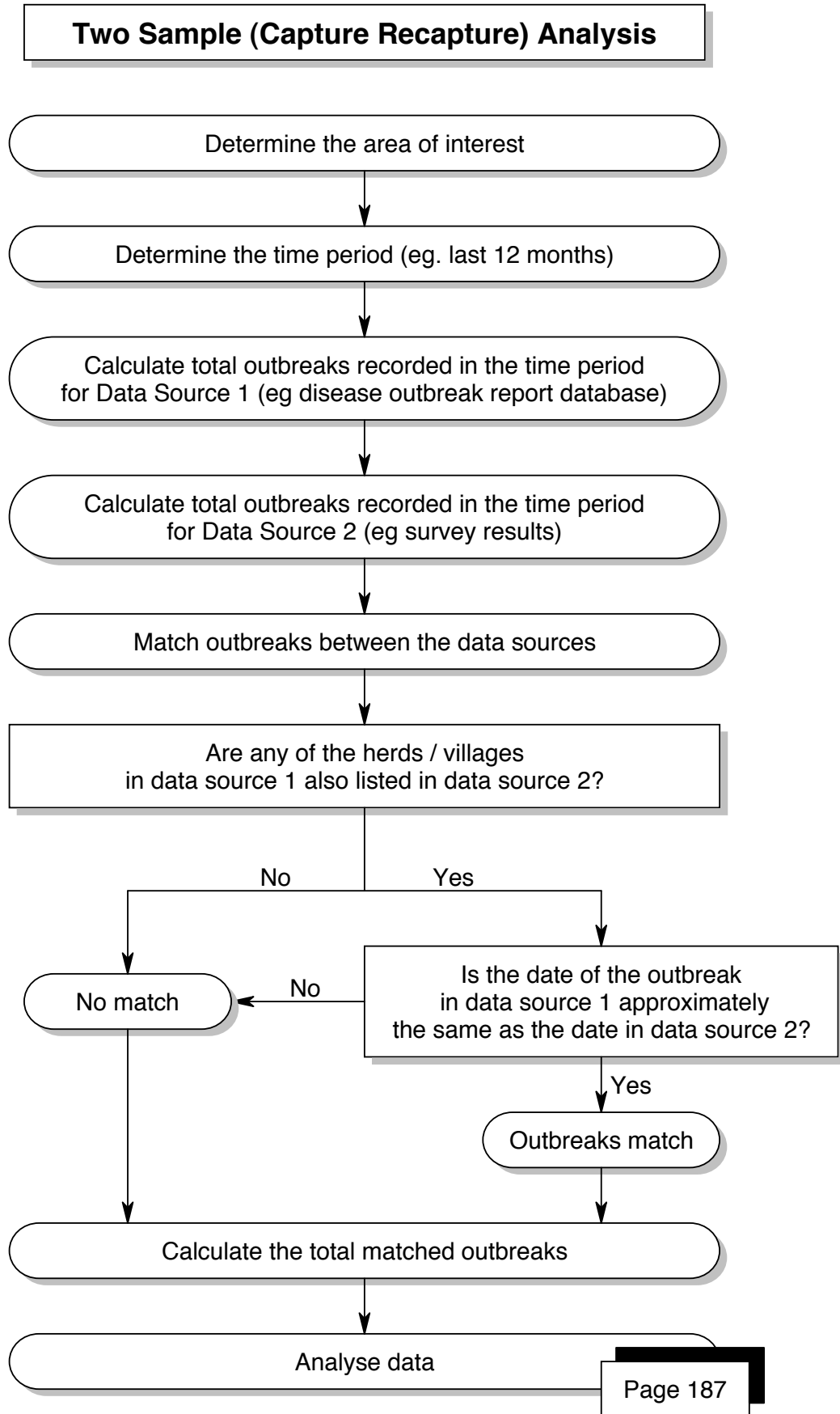
### Background

The technique is known as capture - recapture methodology. It was developed for wildlife studies, where it is very difficult to count every member of a population.

**Example:** A researcher wants to estimate the total number of fish in a lake. It is impossible to count all the fish so a different approach is used. First, for three days, the researcher catches as many fish as possible, and keeps a count of the total. Every fish that is caught is tagged, and then released into the lake. After three days, the fish are left to mix for 2 days. Then the researcher spends another three days catching as many fish as possible. Each time they catch a fish, they record if it has a tag or not. At the end, the researcher has three figures: the total number of fish caught the first time, the total caught the second time, and the total number caught both times (the tagged ones caught a second time).

This is shown in the diagram below. Sample 1 is all the fish caught the first time, sample 2 is the fish caught the second time, and the shaded area is the fish that were caught in both samples.





These figures can be used to estimate the total number of fish in the lake, using a simple formula:

$$Total = \frac{n_A n_B}{n_{AB}}$$

where  $n_A$  is the total in the first sample (A),  $n_B$  is the total in the second sample (B), and  $n_{AB}$  is the total occurring in both sample A and B.

If most of the fish caught the second time already had tags, that would mean that most of the fish in the lake must have been caught already, and the total is only slightly greater than the number of fish caught. On the other hand, if very few of the fish caught the second time had tags, that means that there are many more fish in the lake than were caught the first time, and the total population is much larger.

This same technique can be used to calculate the total number of village or herd disease outbreaks in an area over a specified time interval, and this can be used to calculate the incidence rate. The population is now not fish, but all village or herd disease *outbreaks* (not the villages that had them). The unit of interest is the outbreak (not the village). The first sample is provided by the records of the veterinary authorities of disease outbreaks. These records have 'captured' a certain proportion of all the outbreaks, but probably not all of them. A second source of information is used to 'capture' information about disease outbreaks in the same villages during the same period. The total number of disease records from the first and second sources, and the number of outbreaks that are in both sources can then be used to estimate the total number of outbreaks.

## Data sources

To use the technique, there must be two sources of data on village or herd outbreaks of the disease in question. Furthermore, these two sources must be independent. This means that they should be collected by different mechanisms, and the presence of a particular outbreak in one data source doesn't affect the probability that it will appear in the other.

The first data source usually comes from either disease outbreak reports, or the records of a diagnostic laboratory that is testing specimens from outbreaks. These are both good sources, that should already be available for analysis. The second data source usually comes from a village survey. To be valid, both data sources have to refer to the same time interval, usually one or two years. Only reports or specimens received in a clearly defined period should be analysed, and the survey should record outbreaks occurring only within that same period.

In order to be independent, the same people can't be responsible for collecting both data sources. For instance, the district veterinary officer is usually responsible for submitting outbreak reports. If the second data source comes from a survey in which district officers were asked about village outbreaks, the two sources would not be independent. This is because the chances of an officer remembering an outbreak for the survey are much higher if they have submitted a report on that outbreak.

The best type of second data source is to conduct a survey of a random sample of villages. In general, this can be combined with another survey, for instance a prevalence survey. If another data source already exists (such as the results of an agricultural census in which villages were asked about outbreaks of disease), then this could also be used, avoiding the need for any field data collection.

## Selecting the sample

If a survey is required, the villages or herds should be chosen by simple random sampling from a sampling frame. The **Random Village** program can be used (page 50), using the settings:

- Simple Random Sampling
- Without Replacement
- No stratification

If no sampling frame is available, it is not possible to reliably estimate the incidence rate of village outbreaks. This is because we need to know the total number of villages in order to calculate incidence rate, and without a sampling frame, this is not known.

## Data collection

The data collected are similar to that used in the village outbreak surveys described above. The difference is that instead of just the most recent disease outbreak, the aim is to collect information about the date of every disease outbreak that has occurred over a defined period (usually over the last one or two years). Keeping the time interval relatively short makes it easier for livestock owners to remember. However, if there has been more than one outbreak in the village or herd, livestock owners may find it difficult to reliably recall earlier outbreaks. The techniques described in Chapter 5 (page 118) can be used to improve the quality of data collected during a village interview.

## Data management

Three figures are required for analysis – the total number of outbreaks in the first data source (disease reports or laboratory submission records), the total number of outbreaks in the second source (usually the results of a survey), and the number of outbreaks that appear in both data sources. The first two figures are easily counted. To count the last, outbreaks identified in one source must be matched to outbreaks in the other.

### Matching disease outbreaks

Matching requires good information in both data sources on the village (a village name or identification number) as well as the date of the outbreak. The date reported on a laboratory submission or a disease report will usually not match the date recalled by livestock owners during an interview. This is because there are often small errors in memory, and also because specimens or reports may not be submitted at the beginning of the outbreak. When matching outbreaks, there has to be some flexibility in matching dates. This depends partly on the epidemiology of the disease. For instance, if a disease is very unlikely to occur more than once per year in a single herd, and outbreaks tend to last many months, it may be safe to assume that outbreak dates differing by as much as 6 months or more are, in fact, referring to the same outbreak.

In general, some judgment will be needed for the matching process, and it is best done by hand. Some of the matching may be done with the help of a computer, but even this should be checked by hand.

## Data analysis



Once the totals have been calculated, you can use the **CapRecap** program to analyse the data. Start the program using the Windows Start menu, select Programs, Survey Toolbox, and CapRecap.

- Step 1:** Enter the total number of outbreaks identified in sample 1.
- Step 2:** Enter the total number of outbreaks identified in sample 2.
- Step 3:** Enter the number identified in both samples.
- Step 4:** Click on the **Calculate** button.
- Step 5:** The results show the estimated total, and a 95% confidence interval.

**Capture - Recaputre Analysis**

*Survey Toolbox*

Parameters

Total in sample 1	453
Total in sample 2	762
Total in both sample 1 and 2	232

Results

Estimated population total:	1485.704
95% Confidence Interval:	(1461.612 - 1509.796)

Calculate      Exit

The incidence rate is equal to the total number of outbreaks over the total number of villages multiplied by the time period (page 28). You can use the same formula with the limits of the 95% confidence interval to calculate a 95% confidence interval for the incidence rate.

**Example:** In a two sample study of foot and mouth disease outbreaks over a two-year period, 145 outbreak reports had been received by the veterinary authorities from a province containing 1293 villages. A survey of 85 villages was conducted, of which 47 had experienced outbreaks in the same two-year period. When matched, 36 outbreaks appeared in both data sources. Using the CapRecap program, the estimated total is 188 outbreaks, with a 95% confidence interval from 163 - 213. The estimated incidence rate is therefore 188 outbreaks / (1293 villages × 2 years) = 0.073, or 7.3 outbreaks per 100 villages per year.

# 9

## Surveys to Demonstrate Freedom from Disease

### Contents

Chapter 1: Introduction

#### Part I: Background to Disease Surveys

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### Part II: Survey Design and Analysis

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

Herd or village surveys

Sample size calculation

Selecting the sample

Data analysis

Large area surveys

Sample size calculation

Sample size for minimum cost

First- and second-stage sampling

Data analysis

Other issues

Combining tests

FreeCalc options

#### Part III: Notes for Trainers

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



Consider these three different situations:

- When a disease control program is used amongst intensive farms (e.g. pigs or chickens, or intensive dairy or beef farms) one approach has been to use herd accreditation schemes. These schemes, targeted at farms which supply animals to other producers, involve testing of some farms to provide a guarantee that the herd is free from disease. This means that other producers can buy from that farm without the risk of introducing disease. As a result, the accredited farms are able to ask higher prices for their animals.
- Many national disease control and eradication programs are based on mass vaccination, and later test-and-slaughter. The aim of these programs is to eradicate the disease from the entire country. When this is achieved, all the costs of vaccination and other control measures, as well as losses due to the disease can be saved. However, at the end of an eradication program, the authorities need to be confident that the disease has indeed been eliminated.
- The development of export industries is an important way for developing countries to obtain foreign exchange and develop their economies. The export of livestock or livestock products is one area where agricultural countries have potential to develop an export industry. However, under the rules of the World Trade Organization, an exporting country may be asked to show that there is no risk of spreading livestock diseases to the importing country.

In each of these three situations, it is necessary to demonstrate that a population (either a single herd, village, district, province, or the whole country) does not have a particular disease. This chapter describes survey techniques that can demonstrate that a population is free from disease. In a way, this is the same as conducting a prevalence survey, and hoping that the prevalence is 0, but the theory behind the two types of surveys is quite different.

There are two problems when trying to show that there is no disease in a population. The first is that it is very hard to prove. If a herd has 342 animals, it is only free from disease if none of those animals has the disease. It is possible (although perhaps unlikely) that just a single animal is infected. If we take a sample from the herd and test the animals in the sample, we might, by chance, select the one infected animal, and be able to conclude that the herd is not free from disease. However, it is possible that we might not select that animal in the sample, and think that the herd is free from disease when it is not. The larger the sample size, the higher the chance that we will pick that one infected animal, but there is still a chance that we will miss it. The only way to be completely sure, is to test every animal in the herd. When trying to show that a herd of 342 animals is free from disease, this may be expensive, but not too difficult. When trying to prove that a country with 8 million animals is free from disease, it is impossible to test every animal.

The second problem relates to laboratory tests. In Chapter 2, the concepts of sensitivity and specificity of a test were discussed (page 33). Very few laboratory tests are perfect, and most make a small number of mistakes, calling some diseased animals non-diseased, and vice versa. This means that if you test a large number of animals, it is difficult to interpret the results. If there is one positive test result from 342 animals, is this animal really infected, or is it just that the test has given the wrong result? Are all the animals that tested negative really disease free, or are some of them diseased, and the negative test result was wrong?

A typical test may have a sensitivity of 95% and a specificity of 99%. The sensitivity means that for every 100 diseased animals tested, 95 of them will give a

positive test result, but 5 will give a false negative result. The specificity means that for every 100 disease-free animals, 1 will give false positive test results. This means that even if we test all the animals in a herd or all the animals in the entire country, we still can't be sure if they are all free from disease. Even if there is no disease, we will get some positive test results, because of the test producing false positives. If disease is present, we may also get some false negative test results, and miss the diseased animals.

These two problems mean that it is impossible to prove that a population is free from disease, as there is always the chance that we have missed an animal or that the test result is wrong.

It is impossible to *prove* that a population is free from disease.

Although we can't *prove* that a population is free from disease, if we test enough animals and take the performance of the test (sensitivity and specificity) into account, we are able to show that it is *very unlikely* that the population has infected animals. Surveys to demonstrate freedom from disease do not provide a guarantee, but they are able to say that the chance of the population having diseased animals is smaller than some acceptable level (say 5% or 1%).

If there could be a very small number of infected animals, then it is much harder to find them, and a bigger survey is needed. If the likely number of infected animals is high, then it is easier to find them in a survey, so a smaller sample size is needed. The results of surveys to demonstrate freedom from disease are therefore expressed as the chance that the number of infected animals is equal to or greater than some low value.

**Example:** A outbreak of rinderpest in cattle has occurred in a previously free area. Four neighbouring villages were affected, and the entire cattle population of the villages was slaughtered to eradicate the disease. All villages in the surrounding area (10 km radius) are being examined for clinical signs and serological evidence of infection with the disease. None of the cattle have been vaccinated, so if rinderpest did get into one of these villages, it is likely that it would spread very quickly, and affect a high proportion of the animals in the village, probably over 50%. It is very unlikely that only 10% or fewer animals would be affected by such a contagious disease. During the surveillance after the outbreak, surveys are carried out in each of these surrounding villages, to demonstrate that they are free from the disease. As it is impossible to prove that no animals had been infected at all, the surveys were designed to show that the chance that 10% or more of the animals had been infected with the disease was very low (less than 1%). If fewer than 10% were infected, it was less likely that they would be identified in the survey. Using this survey design all villages were declared free from disease, even though it was possible that some had as many as 10% of animals infected. This didn't pose any risk, because the highly contagious nature of rinderpest meant that if the disease entered a village, much more than 10% of animals would become infected.

Minimum expected prevalence

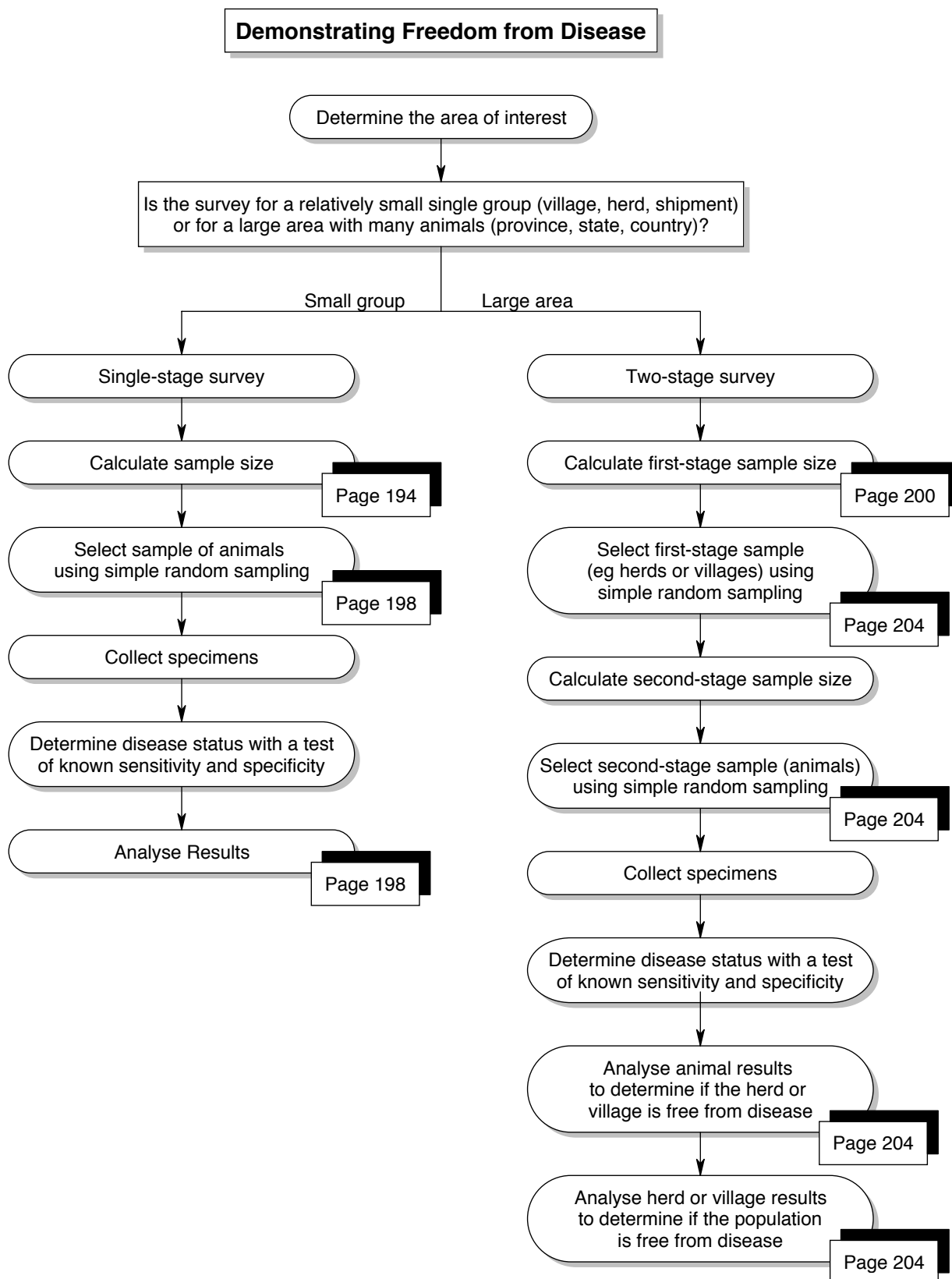
This example demonstrates the concept of the *minimum expected prevalence*. This is the minimum prevalence expected if a contagious disease was present in a herd. When conducting a survey, this is the lowest disease prevalence that the survey can be expected to reliably identify. If the disease is present in the population, but at a lower level, than the prevalence specified, then the survey may not be able to identify it. This level is based on a knowledge of the epidemic behaviour of the disease. For example, FMD may have a minimum expected prevalence of 30%, while Johnes disease may have a minimum expected prevalence of 5% or less.

Maximum acceptable prevalence

For diseases which are not highly contagious, the minimum expected prevalence may also be thought of as the *maximum acceptable prevalence*. If the level of disease in the population is less than this prevalence, then it is small enough not to worry about. This level always has to be greater than 0, because unless we have a perfect test and are able to test every single animal, we can't prove that the prevalence is 0.

As with any survey, it is easier to find diseased animals if the prevalence is higher. The measure of disease should therefore be that measure which gives the highest prevalence. As discussed in Chapter 2 (page 31), antibody status lasts much longer than clinical disease, and therefore has a higher prevalence. This is why it is much more common to base a survey to demonstrate freedom from disease on the use of serological tests rather than clinical examination of animals.

This chapter describes two different survey designs for demonstrating freedom from disease. The first is a survey conducted in a small population, such as a herd, village, or intensive farm. The second design uses a two-stage sampling scheme for surveys of larger areas (districts, provinces, states or countries).



## Herd or village surveys

The main steps in conducting a single-stage survey to demonstrate freedom from disease are:

- Step 1:** Determine what question is being asked. This involves specifying the minimum expected prevalence (maximum acceptable prevalence), and the probability levels which determine how confident we are of the results.
- Step 2:** Calculate the sample size.
- Step 3:** Select the sample using simple random sampling.
- Step 4:** Collect specimens.
- Step 5:** Process specimens ready for analysis.
- Step 6:** Send the specimens to the laboratory.
- Step 7:** Check the data for completeness and accuracy.
- Step 8:** Analyse the data to determine the probability that, if disease is present in the population, the prevalence is less than the maximum acceptable prevalence.
- Step 9:** Report the data.

## Sample size calculation

Calculation of the sample size for a survey to demonstrate freedom from disease is based on several different values.

### Test performance

The performance of the test being used plays an important role in the sample size. It is expressed in terms of sensitivity and specificity (page 33). If the test is not very reliable (either sensitivity or specificity or both are relatively low) then the sample size will need to be much higher. If there is a choice, the test with the best sensitivity and specificity (but particularly high specificity) should be used. See Combining Tests (page 205) for advice on ways to improve the specificity of a test.

Unfortunately, precise estimates for sensitivity and specificity are not always available for many tests. Another problem is that these measures vary somewhat depending on the population being tested, so that values published from a study in one part of the world may not be completely valid for the population being surveyed. If you don't know what the sensitivity and specificity of tests are, here is what you should try:

- Step 1:** Ask your laboratory people if they have conducted any studies in the local population to evaluate test performance.
- Step 2:** Ask them if they know of published figures based on other populations.
- Step 3:** Search the literature for published studies on the test. If more than one study is found, use the one that most closely matches your population.
- Step 4:** Contact leading experts with experience in using the test, and ask them for their estimates of the test performance.
- Step 5:** Organise a small study in the local population to measure test performance yourself.

If no reliable published figures are available, then estimates may be used. However, if the test is going to form the basis of important surveys, or be used as part of an ongoing control or eradication program, it is very important that its performance in the local population is well understood. It may be worthwhile to conduct a study to evaluate the performance of the test. This involves testing a number of animals that have a known status (some truly disease positive, and some truly disease negative), and directly calculating the sensitivity and specificity. Consult an epidemiology text or epidemiologist for advice on conducting this sort of study. Also, when finished, make sure that you publish the results, so that others can benefit from your work.

### Population size

You need to know the size of the population. Smaller populations require somewhat smaller sample sizes.

### Minimum expected (maximum acceptable) prevalence

The choice of this figure (explained earlier) is based on a knowledge of the disease, or practical limitations. The larger the prevalence chosen, the smaller the sample size and the easier the survey. With highly contagious diseases that are likely to spread quickly, it is safe to choose quite a high prevalence. However, with other diseases, the prevalence in a herd, if present, may be very low. Detecting a disease at low prevalence can be very difficult, requiring a large sample size. In the end, you may have to be content with a prevalence level that is based on the largest sample that can be afforded or practically tested, rather than on the biology of the disease.

### Type I and II error

Type I error

The *Type I error* (also called the  $\alpha$  (alpha) level) is the probability that the results of the survey will conclude that the population is not diseased when in fact it is. This is also known as the significance of the results, and is equal to 1 minus the level of confidence. The Type II error ( $\beta$ , beta) is the probability that the survey will conclude that the population is diseased, when in fact it is not. This is equal to 1 minus power. By convention, the Type I error is usually 0.05, and the Type II error is either 0.1 or 0.05. These can be adjusted to any value, depending on the importance of that type of error.

Type II error

**Example:** A piggery is being tested as part of a herd accreditation scheme. If the herd is found to have disease, then the owner will not be allowed to sell animals other than for slaughter. The owner is therefore very keen to make sure that the survey doesn't make a type II error, or conclude that the farm is infected when it is not infected. The owner would want to set the type II error level to be very low, to minimise the chance of this mistake. On the other hand, a client of the farm that buys pigs for breeding does not want to receive diseased pigs. The client would want to make sure that if the survey indicated that the farm was free from disease, this is in fact true. The client would want a very small type I error level to ensure that the farm is not declared free when it is actually infected.

The final decision on the error levels depends on a compromise between competing needs. The repercussions of the possible mistakes need to be taken into account as well, as in the following example.

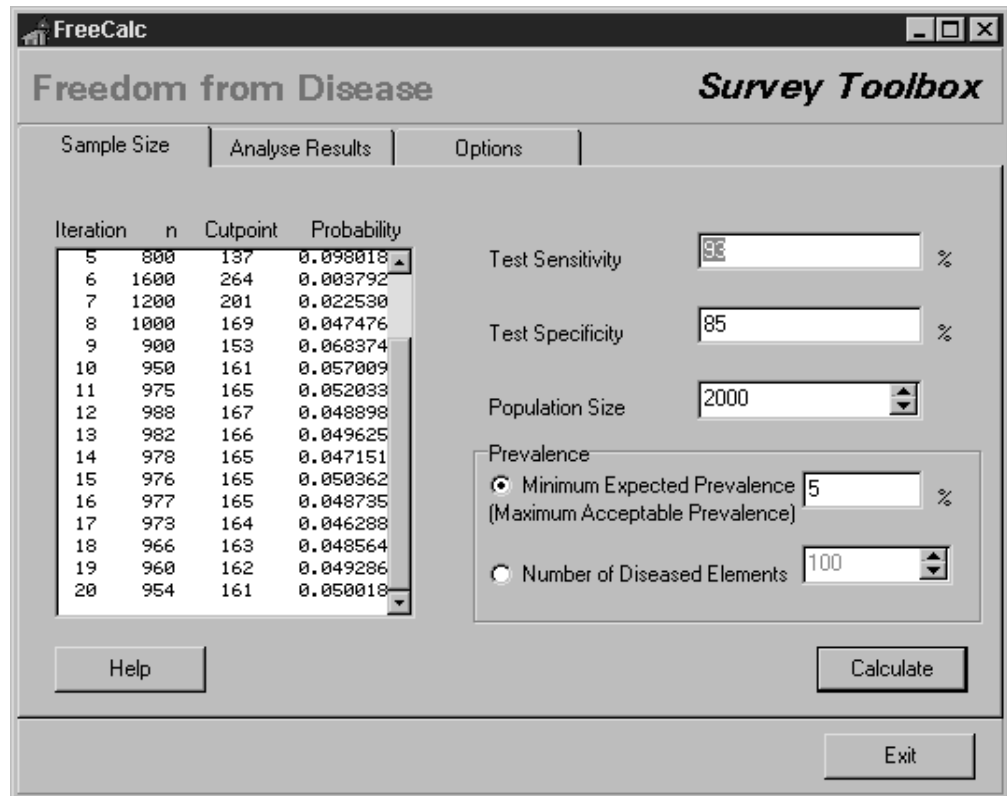
**Example:** After an outbreak of FMD, villages near the outbreak are being monitored. A survey is conducted in each village to determine whether it is free from disease. If it is free, quarantine is removed, and the village can trade again. If it is not shown to be serologically free, it is kept under quarantine. (If clinical cases are detected, the population of the village is slaughtered.) If the survey makes a type I error, and concludes that the village is free from disease when it does have disease, the consequences could be very bad. The disease could spread from that village to other parts of the country and the outbreak could start again, causing the death of many animals, and enormous expense. The probability of a type I error should be kept very very low. If a type II error is made, then the village is held under quarantine for a bit longer. This is inconvenient for the livestock owners in that village, but doesn't have a huge impact. The type II error probability could therefore be quite high.

### Calculating the sample size

When all these issues have been considered, you can use the **FreeCalc** program to determine what sample size is needed. Use the Windows start menu, select Programs, Survey Toolbox, then FreeCalc. FreeCalc is a program both for sample size calculation, and analysis of survey results.

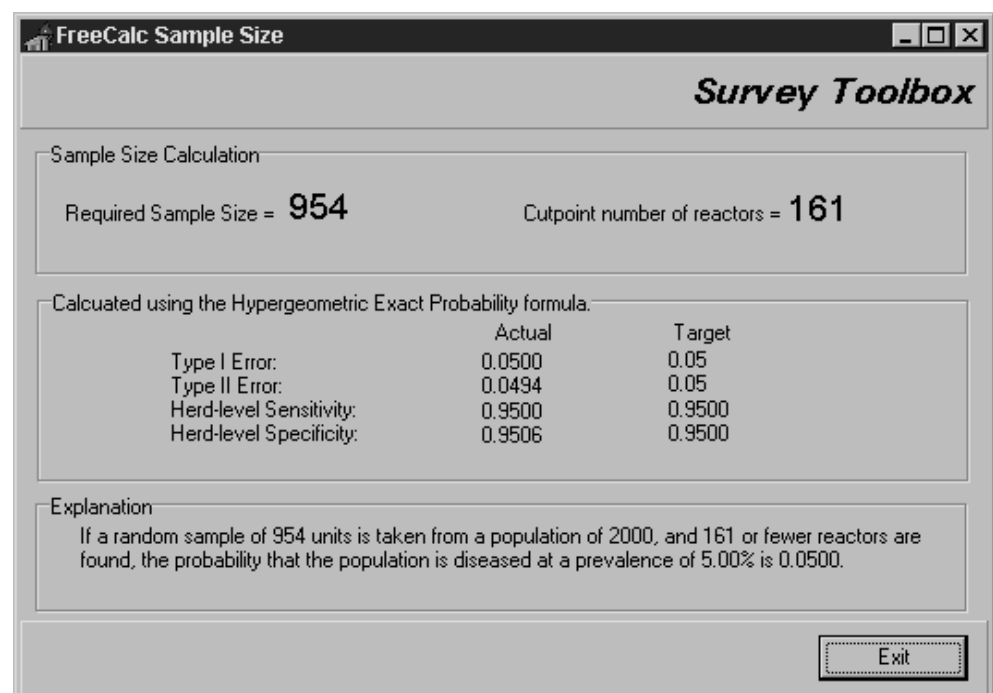


- Step 1:** Click on the Sample Size tab at the top of the window.
- Step 2:** Enter the test sensitivity and specificity as a percentage.
- Step 3:** Enter the size of the total population in the Population Size box.
- Step 4:** In the Prevalence box, enter the minimum expected prevalence (the maximum acceptable prevalence). This can either be entered as an estimate of the prevalence (a percentage) or as a direct measure of the number of diseased animals in the population. Click on the ratio button to choose if you want to enter the prevalence or the number of diseased animals, then type in the number. The equivalent value will be shown in the other box.
- Step 5:** Click on the Options tab at the top of the window.
- Step 6:** In the Formula for Calculation box on the left, you can usually leave this as Modified Hypergeometric Exact. The different formulae are discussed on page 206.
- Step 7:** In the Parameters box, enter the Type I and Type II error levels that you want to use. If you are unsure, leave them both at 0.05. (The other values in the box are discussed on page 207.)
- Step 8:** Click on the Sample Size tab again, and click on the **Calculate** button. As the calculation is taking place and the program is searching for the best sample size, the intermediate results are displayed in the box on the left.
- Step 9:** When finished, the results are displayed in a window.



The results show the sample size required to be confident that, if the disease is present, it is present at a level lower than that specified for maximum acceptable prevalence.

The results also show the ‘cutpoint number of reactors’. This is the number of animals that can return positive test results, and still let us conclude that the herd is free from disease. In other words, these are considered to be false positive test results. If we get fewer test-positive animals in the survey, we can still conclude that the herd is free from disease, but if there are more than this number, the evidence for being free from disease is not as strong.



## Selecting the sample

When the sample size has been calculated, you are ready to conduct the survey. Animals must be chosen using simple random sampling. If a sampling frame already exists, you can use the manual technique described on page 41, or, if it is available on disk, you can use the **Random Village** program (page 50). If no sampling frame exists, one must be made. If conducting a survey in a village, use the technique described for building a livestock sampling frame on page 115. If on a farm, it may be possible to select a systematic random sample, as long as there are some livestock handling facilities (page 44). See Chapter 3 for a full discussion of random sampling.

## Data analysis

When the sample has been examined, or specimens analysed in the laboratory, the key pieces of information required are the total number tested (the sample size), and the number that gave positive test results. You can then use the FreeCalc program to analyse the results



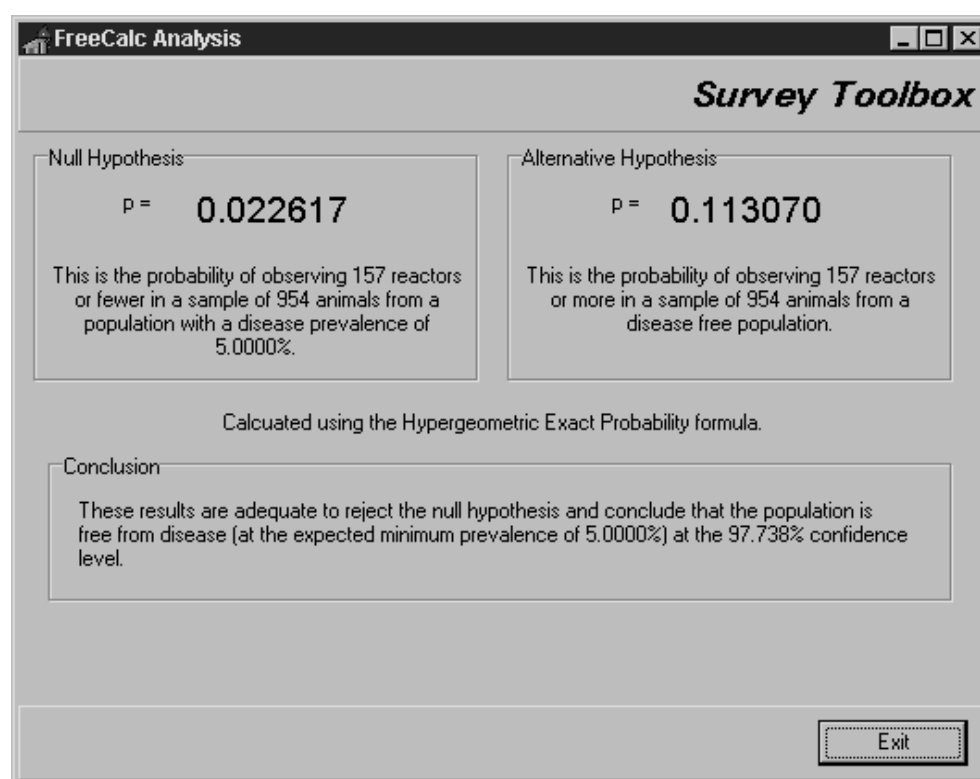
- Step 1:** Start the FreeCalc program (as described earlier) and click on the Analyse Results tab at the top of the window.
- Step 2:** Enter the test sensitivity, specificity, population size and prevalence as described for sample size calculation.
- Step 3:** On the left, enter the Survey Sample Size.
- Step 4:** Enter the Number of Positive Reactors from the test results.
- Step 5:** Click on the Options tab, and check the Type I and Type II error levels, to make sure they are correct.
- Step 6:** Return to the Analyse Results tab, and click on the **Calculate** button.
- Step 7:** A window is displayed with the results of the analysis.

The screenshot shows the 'FreeCalc' window titled 'Freedom from Disease Survey Toolbox'. The 'Analyse Results' tab is selected. The interface includes the following elements:

- Survey Sample Size:** A dropdown menu showing the value 954.
- Number of positive reactors:** A dropdown menu showing the value 157.
- Test Sensitivity:** A text input field containing 93, followed by a percentage sign (%).
- Test Specificity:** A text input field containing 85, followed by a percentage sign (%).
- Population Size:** A dropdown menu showing the value 2000.
- Prevalence:** A section with two radio buttons:
  - Minimum Expected Prevalence: 5 %
  - Number of Diseased Animals: 100
- Buttons:** 'Help', 'Calculate', and 'Exit' buttons are located at the bottom of the window.

The results are displayed in terms of probabilities of the null and alternate hypothesis. The probability of the null hypothesis is the probability of observing this many reactors or fewer, if the population was diseased at a level equal to or greater than the specified prevalence. If this probability is small, we can conclude that it is very unlikely that the population is diseased. If the probability is large, then there is not enough evidence to conclude that the population is free from disease.

The probability of the alternative hypothesis is also shown. If this is small, then it is very unlikely that the population is free from disease. If it is large, then it is consistent with there being no disease in the population. If both the null and alternative probabilities are small, it suggests that the population is not free from disease, but the prevalence is less than the minimum expected prevalence specified. The conclusion is written at the bottom of the window.



## Large-area surveys

In Chapter 3, the problem of drawing a simple random sample from a large population was discussed. For surveys at the district, state, or national level, it is not possible to draw up a sampling frame which lists every animal in the entire population. In these cases, two-stage sampling (page 64) is much more practical. At the first stage, we need only a sampling frame that has all the herds or villages listed. For those herds or villages that are chosen, we can then build a sampling frame of animals.

The same approach is used for surveys of large areas to demonstrate freedom from disease. Two-stage sampling has the added advantage that it is able to account for *disease clustering*.

Disease clustering

Disease is not usually spread evenly through the population, but tends to occur in clumps or clusters. For instance, with Newcastle disease in chickens, most of the

animals in most farms and villages are not affected, and the overall prevalence in the population is very low. However, during an outbreak, a small number of villages or farms have a large number of birds affected, and the prevalence in those farms or villages is very high.

Two-stage sampling allows us to account for the fact that if a disease is present, very few villages or herds may be affected, but those that are usually have relatively high levels of disease. This is taken into account by specifying prevalence at two levels, the prevalence of infected farms, herds, or villages at the first level, and the prevalence of infected animals on farms at the second.

The procedure for conducting a two-stage survey to demonstrate the freedom of a large area from disease is as follows:

- Step 1:** Determine what question is being asked. This involves specifying the minimum expected prevalence (maximum acceptable prevalence) both amongst herds or villages and amongst animals within a herd or village.
- Step 2:** Calculate the first stage sample size (number of herds or villages).
- Step 3:** Select the sample using simple random sampling
- Step 4:** Build a sampling frame of animals within the selected herds or villages
- Step 5:** Calculate the sample size depending on the village or herd population.
- Step 6:** Select animals using simple random sampling.
- Step 7:** Collect specimens.
- Step 8:** Process specimens ready for analysis.
- Step 9:** Send the specimens to the laboratory.
- Step 10:** Check the data for completeness and accuracy.
- Step 11:** Analyse the data from each village or herd, and determine whether it is classified as diseased or non-diseased.
- Step 12:** When every village or herd in the sample has been classified, analyse these herd-level results to determine if the entire population is diseased or non-diseased.
- Step 13:** Report the data.

## Sample size calculation

The sample size calculation has two steps - first calculate the number of herds or villages required, and then the number of animals from each herd or village.

In addition to all the measures required for sample size calculation for small populations, there is another important measure to be considered. For each herd that is tested, we need to decide whether the herd is to be classified as diseased or non-diseased. We analyse the results from animals within the herd to make this decision. However, this decision might be wrong. The probability of making a wrong decision is given by the Type I and Type II error levels.

When making a decision (or diagnosis) about an entire herd or village, the procedure (testing a sample of animals from the village) can be thought of as a test. This is often called *herd testing*. Just like any other test, its performance can be measured by sensitivity and specificity. The sensitivity of a herd test is the probability that a diseased herd will be classified as diseased. This is equal to 1 - Type I error level. If, when testing animals within a herd, we set the Type I error level

to 0.05 or 5%, then the sensitivity of the herd test is 0.95 or 95%. In the same way, the specificity of the herd test is equal to  $1 - \text{Type II error level}$ .

When we set the Type I and II error levels for determining the sample size within a single herd, we are in fact setting the sensitivity and specificity of the 'test' for that herd. With this in mind, we can go ahead and determine the sample sizes needed for two-stage sampling.

To calculate the number of villages or herds that need to be selected at the first stage, use the **FreeCalc** program:



- Step 1:** Click on the Sample Size tab.
- Step 2:** Under test sensitivity, enter the a value which is  $1 - \text{Type I error}$  used for selecting individual animals. For instance, if a Type I error level of 0.05 is used when selecting individual animals to test, this means that the herd test sensitivity is 95%.
- Step 3:** Enter the specificity in the same way. If the Type II error level for selecting individual animals is 0.1, then the herd-level specificity is 90%.
- Step 4:** Enter the size of the total population. This is the total number of herds or villages in the area being studied (not the total number of animals).
- Step 5:** In the prevalence box, enter either the prevalence or total number of disease-positive villages or herds which represents the maximum acceptable prevalence. Regardless of the disease in question, if the population is thought to be free from disease, then the proportion of positive villages must be set to a relatively low value (usually less than 5%). This means that the number of villages or herds that need to be tested will often be quite high.
- Step 6:** Click on the Options tab, and check the Type I and II error levels. These now measure the probability that the entire survey will make an error. See the discussion on page 195.
- Step 7:** Return to the Sample Size tab, and click **Calculate**.

The results will indicate how many villages or herds need to be visited. The procedure for selecting animals from each of the selected villages is the same as that described above for surveys of small populations on page 194. At this level, the sensitivity and specificity are measuring the performance of the laboratory test. The Type I and II error levels, which determine the herd test sensitivity and specificity, are set to the values mentioned in steps 2 and 3 above. The population size refers to the total number of animals in that village or herd. Because the sample size changes depending on the total population, and the population of every village is likely to be different, the second-stage sample size should be calculated for each village separately. If a portable computer is not available, the sample sizes for every possible village size can be calculated before the fieldwork, and written in a table for use in the field.

## Sample size for minimum cost

When using two-stage sampling, a survey can produce a result of the same accuracy by using a variety of combinations of first and second-stage sample sizes. For instance, if a small number of villages is selected and many animals from each village are tested, it is possible to get the same accuracy as if many villages are tested,

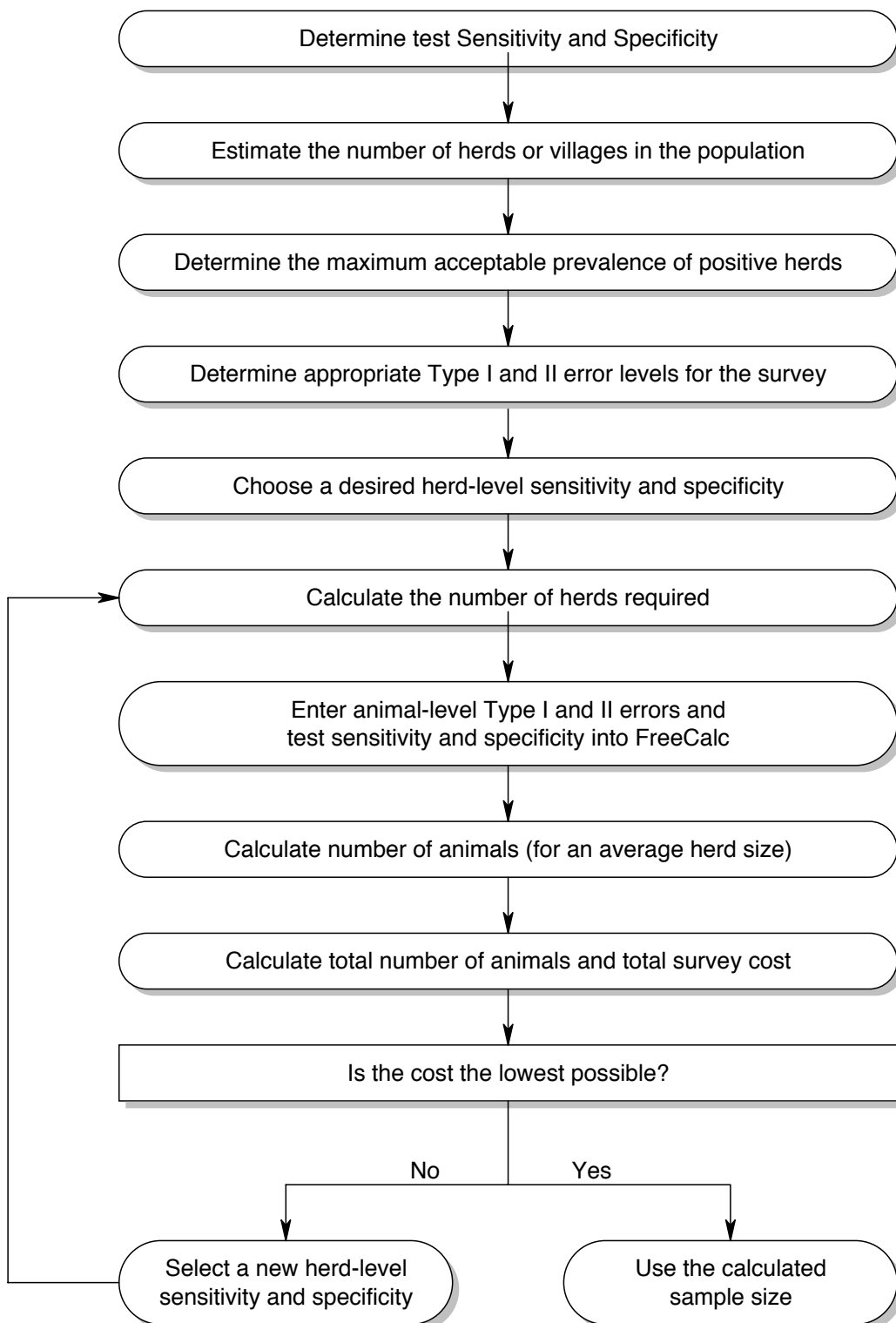
and only a small number of animals are tested from each village. By changing the Type I and II error levels used for selecting the sample size for the second stage (testing animals within a village or herd), we are also changing the sensitivity and specificity of the herd test (used when selecting villages or herds at the first stage). This enables us to produce a variety of different sample size combinations, all of which will provide the same level of evidence for freedom from disease.

This flexibility is one of the advantages of two-stage sampling, because not all the combinations will cost the same. The overall cost depends on how much it costs to test a single animal, and how much it costs to test a single herd or village. This was discussed in Chapter 7 (page 154). For prevalence surveys, there is a formula, used in the Prevalence program, to determine what is the cheapest combination. For surveys to demonstrate freedom from disease, the complexity of the calculations means that it is not possible to use a formula to work out the best combination.

Instead, it can be done using trial and error with the FreeCalc program. Use the following procedure to calculate the best combination of first and second-stage sample sizes:

- Step 1:** Determine the basic measures that we can't change. This includes the sensitivity and specificity of the laboratory test, the population of villages or herds (first-stage population size), an estimate of the average animal population of the villages or herds, the maximum acceptable prevalence of the disease amongst villages (first stage) and animals (second stage), and the overall Type I and Type II error levels for the survey (used when calculating first-stage sample size). You also need to know the cost of testing a single animals, and the costs associated with sampling a single village (see page 154).
- Step 2:** Pick starting values for the herd test sensitivity and specificity. The higher these values are, the fewer villages need to be tested, and the more animals need to be tested in each village. If they are very high, there may not be enough animals in some villages to achieve this level. In general, try to make the specificity as high as possible.
- Step 3:** Calculate the number of herds needed using the selected herd test sensitivity and specificity.
- Step 4:** Now use the same figures to calculate the second stage sample size. Set the Type I error to  $1 - \text{Sensitivity}$ , and the type II error to  $1 - \text{Specificity}$ . Change the sensitivity and specificity to those of the laboratory test, the population to the average herd or village size, and the prevalence to the maximum acceptable or minimum expected prevalence within the herd.
- Step 5:** Calculate the number of animals that need to be tested.
- Step 6:** Using the number of herds and villages, and the number of animals, calculate by hand the total cost of the survey, based on the cost estimates, and record the result.
- Step 7:** Now go back to calculating the first-stage sample size, but change either the sensitivity or specificity or both. Repeat the calculations in steps 3 to 6, and record the sample sizes and total cost of this alternative combination.
- Step 8:** Continue testing new values until you find the one that gives the cheapest cost.

**Calculating Optimal Sample Sizes for Two-Stage Surveys to Demonstrate Freedom from Disease**



## First- and second-stage sampling

At the first stage, villages or herds must be selected using simple random sampling from a sampling frame. Use a random number table to select from a written sampling frame or the **Random Village** program (described on page 50) to select from a sampling frame on computer disk. When using the program, set the Sampling Type to Simple Random, and select villages or herds without replacement. The sample may be stratified if convenient.

At the second stage, animals must again be chosen using simple random sampling, or, if possible, systematic random sampling. For selecting animals in a village, the technique described in Chapter 3 (page 55) can be used, either with the Random Animal program, or manually using a random number table. If a computerised sampling frame of all animals in a herd is available, you can again use the Random Village program, with the same settings.

## Data analysis

The data are analysed in two stages. First, the data from each selected herd or village are analysed to provide a herd result, indicating that the herd is either diseased or non-diseased. Use the same approach described on page 198 for small population surveys, and record the status of each herd or village. When analysing the results from each herd, be sure to enter the correct population size for that herd, and the correct Type I and II error levels selected for the second stage of sampling. The sensitivity and specificity should be those of the laboratory test.

When all herds or villages have been analysed separately, the population of herds or villages can be analysed. Use the FreeCalc program:

- Step 1:** Start FreeCalc and click on the Analyse Results tab.
- Step 2:** Enter the herd test sensitivity and specificity.
- Step 3:** Enter the total number of herds or villages for the Population Size.
- Step 4:** Enter the maximum acceptable prevalence in the Prevalence box.
- Step 5:** Check the Type I and Type II error levels in the Options tab. They should be the error levels for the overall survey, not for the second stage of testing.
- Step 6:** Return to the Analyse Results tab, and enter the Survey Sample Size. This is the total number of herds or villages that were tested (the first-stage sample size).
- Step 7:** In the Number of Positive Reactors box, enter the total number of herds that were classified as diseased.
- Step 8:** Click the **Calculate** button.

Although some of the herds were classified as being disease positive, the herd test was not perfect and may have made a small number of mistakes. This final stage of analysis calculates whether the number of positive herds can be accounted for by the errors in the herd test. If so, then it is possible to conclude that the population is free from disease. If the number of positive herds is too high, then the population must still be classified as diseased.

## Other issues

### Combining tests

Often the performance (sensitivity and specificity) of a laboratory test is not as good as we would like it to be. If the specificity in particular is low, then the sample sizes to demonstrate freedom from disease will be very high. Sometimes it might not be possible to achieve the confidence required by testing every animal in the population. One way to address this problem is to test each specimen with two different tests. Depending on the way the results are interpreted, this can dramatically improve either sensitivity or specificity (but not both).

**Example:** During the last stages of a tuberculosis eradication campaign, the specificity of tuberculin skin tests, although high, is not perfect, and when testing very large numbers of animals may produce many false positive reactors. One approach is to test animals that produce a positive tuberculin skin test result with another test (e.g. gamma interferon). If the second test is also positive, then the animal is considered infected, but if the second test is not, then the animal is considered uninfected. This type of testing increases the specificity of a test, but decreases the sensitivity (and therefore the proportion of false negatives).

#### Tests in series

To increase the specificity of a test, two tests can be used in series. The first test is applied, then, if the animal is positive, a second test is used. The animal is considered positive only if both tests are positive. Animals that test negative to the first test are not retested. The result of this approach is to increase the specificity, but decrease the sensitivity. The overall sensitivity ( $Se_t$ ) and specificity ( $Sp_t$ ) of the two tests combined (the "test system") will be:

$$Se_t = Se_1 \times Se_2$$

$$Sp_t = Sp_1 + Sp_2 - (Sp_1 \times Sp_2)$$

Tests can be used in series in a similar way to increase sensitivity. If an animal tests positive with a single test, then it is considered positive, but if it is negative, it is retested, and considered negative only if it produces a negative result in the second test as well. In this case the specificity decreases:

$$Sp_t = Sp_1 \times Sp_2$$

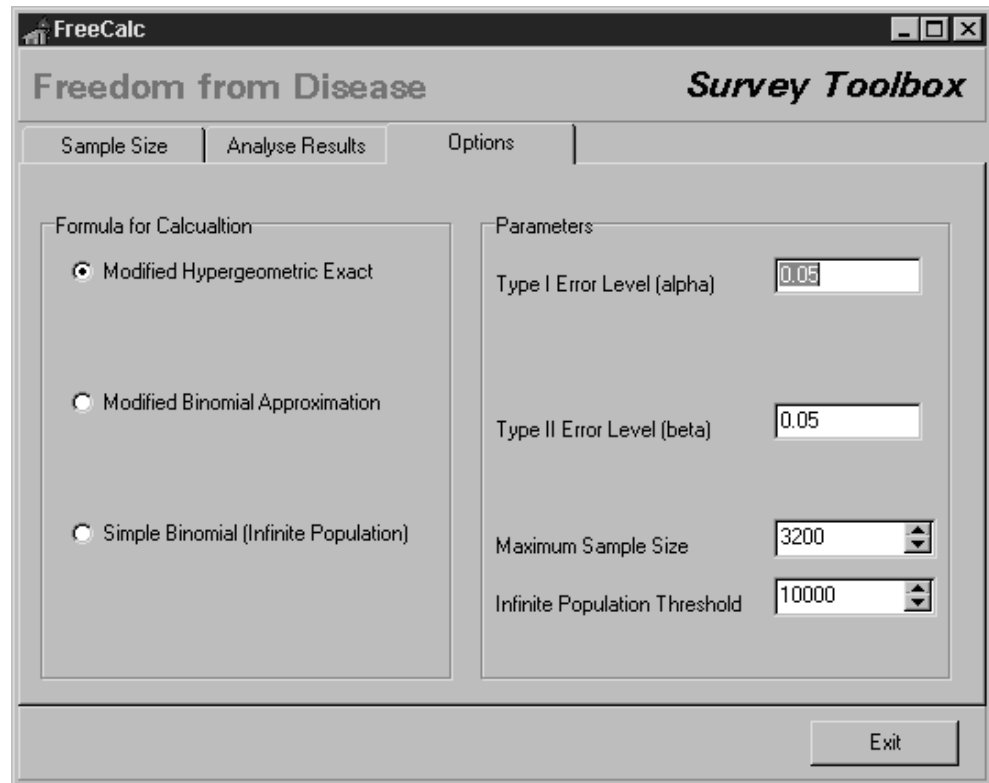
$$Se_t = Se_1 + Se_2 - (Se_1 \times Se_2)$$

#### Tests in parallel

Another approach is to test every sample with two tests, and base the final result on both the results. If both tests are positive or both are negative, the result is clear. However, when the two tests disagree, a decision must be made as to what the result is. If animals with conflicting results are considered negative, the sensitivity decreases and the specificity increases, as demonstrated by the equations in the first example above.

If conflicting results are interpreted as positive, then the equations in the second example above should be used, with a decrease in specificity and an increase in sensitivity.

## FreeCalc options



### Formulae

On the options tab of the FreeCalc program, there is a choice of three different formulae to use for the calculations.

The first is the Modified Hypergeometric Exact Formula. This formula calculates the exact probabilities for sample sizes and analysis of results. Under certain circumstances, this formula requires an enormous number of calculations, and can therefore be very slow. This happens when the sample size is large, due to poor test performance (especially low specificity) or a small maximum acceptable prevalence. Nevertheless, you should use this formula all the time, unless you find that calculations are becoming too slow.

The Modified Binomial Approximation Formula calculates the same probabilities, but uses an approximation, making the calculation faster. This formula still produces accurate results except when the sample size very large, relative to the population size. Use this formula if the Modified Hypergeometric Exact formula is too slow, and the sample size is less than half of the population size. Although much faster than the Exact formula, for very complex calculations (very large sample sizes) this formula can also become quite slow to calculate.

The Infinite Population Binomial Formula is the fastest to calculate. It assumes that the size of the population is infinite (or at least very much larger than the sample size). If you are working with very large populations, and the other two formulae are slow to calculate, use this formula. When the population is not very large, the use of this formula can lead to significant errors.

**Maximum Sample Size**

You can specify the maximum sample size for the program to calculate. If the sample size required is larger than this maximum, the program displays an error message, and stops the calculation.

**Infinite Population Size**

When the population is larger than a specified size, the program automatically uses the Infinite Population Binomial Formula, regardless of which formula has been selected. With very large population sizes, there is virtually no difference between the Exact and the Infinite Population formulae, so the faster of the two is used. You can enter the size of the population above which the faster formula will be used.

o o

ar n n n r a n n r nar  
a n r an n na na rr na an  
n rna na n a a r n r r a n n a r r  
an r n an n r r a n r n a r  
a ar n a ar an ar n an n r an a r a  
r a r r n an r a r r a n n r  
r n a a r a n a r a a  
an r n a r a n a r n a r ar an r r  
n an a r a r a r an a n r

o

n an n a r a n r a a r a  
n r a r n a n n a a an r  
r a n r a r n n an a a n n  
ar ar r r a n n ar n r an  
a a n n r n an na n r a  
r a arr a ar n r n a  
r n n r n a a n r r n r n  
n r a n r r n n an n r r  
a r r a  
an n r n a r a n n a a ar n  
n a a r ar r r a r a a a an  
r r n r a n n r r a n n an r r ar n an  
a a ar a r r r a n n r ara r a n n  
r

a r

r r a n r

o

r r r nar a ar r n r ann n  
 r an n an r nn n r ar a r n r a a ana n  
 an ana a an na na r r n a r nar a  
 r r n a a a r n n r an n n ar  
 r n a r ar am n a n r ara n  
 n arr ara r

o

r n r a r n a n r r  
 a r a ar r n r aaan n n n  
 r r a a a r a r r n a n  
 r a na n a r r ra r an ar r  
 r an a na ran r r na r a a a  
 n r an n a an n n r a r n  
 r an a a ar an r r a ar a  
 n n  
 n n r r r a r ar n  
 a a ar an a an a r n ra a  
 arn

o

n r n a r r a  
 ar an a r r r na r ar a r  
 ana n an ana r a a n a a a a n  
 ran r r n n r a n r r a r n r an  
 n r r a n r a n r r

o

r r r n an n r a n an ra n r  
 n n ran ar a ar r an an  
 a an n r r  
 n nn r n r r n rn n n  
 n ra a r a a n ar an a r a ar r an  
 r ar an ar an ann a r a  
 n r r

# 10

## Guide for Trainers

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Advice for trainers

Who should be a trainer

Training skills

Lesson planning

Activities

Teaching Techniques

Learning land marks

Reinforcement and practice

Warmers

Questions and answers

Games or competitions

Group discussions

Brainstorming

Ranking

Role playing

Field trips

Practical activities

Field work

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



## Advice for trainers

It is often assumed that anybody who understands a subject well should be able to teach that subject to others. Unfortunately, this is not the case. There is a lot more to good teaching than just having an understanding of the subject. An understanding of the students and the way they learn is also necessary.

This chapter discusses who should be a trainer of the active surveillance techniques described in this book, and provides advice on techniques that may be used to help with the training.

## Who should be a trainer

The two basic requirements for a trainer of active surveillance techniques are a good understanding of the subject and the ability to teach it to others. The most likely people to be involved in training are national or provincial level veterinary staff (epidemiologists) responsible for livestock disease control. Veterinary epidemiologists working as development project staff may also be involved. Other people with different backgrounds may also successfully conduct training courses, but ideally, the trainer should have the following characteristics:

- A sound understanding of active surveillance, and the techniques described in this book. Experience with surveillance, and sample surveys, and a knowledge of epidemiological principles is important. However, in the absence of formal epidemiological training, a sound knowledge of all the concepts in this book, and experience in conducting survey field work will provide a trainer with all the technical background necessary.
- Practical field experience. The trainer should be reasonably experienced in restraining livestock and collecting specimens.
- Ability to use computers. Many of the technical calculations and analyses depend on the use of computer programs. The trainer should be familiar with computers and the programs being used (including a database program such as Epi Info), and be able to solve the types of computer problems that may arise.
- An ability to communicate easily with trainees. The trainer should be reasonably fluent in their language, and have an understanding of the social and cultural issues that may impact on field work and training.
- A respect for the skills and experience of trainees, and the knowledge of livestock owners.
- Enthusiasm for teaching and for active surveillance and survey field work. Enthusiasm is contagious.
- Experience with, or an understanding of basic teaching techniques. These are discussed in more detail below.

## Training skills

Every trainer has their own style. While you may wish to copy some good points from people you have learnt from, there is no point trying to imitate them completely. Some people are more serious and strict, while others are casual and like to joke a lot. Both can make a good trainer, as long as they are comfortable with the way they do things. Whatever your style, try to consider if anything you do makes it more difficult for students to learn. If so, then try to change it. Whatever your

training style, it is always possible to learn new tricks and techniques, and to improve the effectiveness of your training.

Training is a process of communication, both from the trainer to the students, and from student to trainer. There are many ways to make this communication more effective. You are not simply transferring information that you have to others, but you are trying to help the participants understand and solve problems, using tools they already have and new tools you give them. Some tips to encourage effective communication include:

- Keep eye contact with the participants. Don't talk with your back to them, while you are working on the board.
- Show interest in what you are saying, and make it more like a story. Don't speak in a droning voice. Speak clearly and loudly, but not too fast.
- Vary activities regularly, so participants don't become bored. Don't spend too much time in the classroom.
- Make sure that the environment where you are teaching is comfortable and not too distracting.

## Lesson planning

One of the keys to successful teaching is good organisation and planning. Regardless of how much technical knowledge the trainer has, if they are not organised, or are unsure what they are doing next, then the students will find it difficult to learn.

A well organised, carefully thought out lesson plan will ensure that both the students and the trainer know exactly what is happening, and that effective learning can take place. Lesson plans have been prepared for the training courses suggested, and are presented in Chapter 11. These should be thought of only as a guide, as the specific needs of each training course will be different. You may use some of these lesson plans if they are appropriate, or develop your own. The structure that has been used for the lesson plans in this book is as follows:

- Title. The title of the lesson, so students know what to expect.
- Location. Where the lesson is to be conducted (classroom, village etc.)
- Duration. The expected time of the lesson. This can vary greatly depending on the level of knowledge and experience of the students.
- Objectives. These are the things that the student should be able to do at the end of the lesson.
- Key points. These are highlights from the lesson, and things to keep in mind when teaching.
- Page references. The relevant pages from this book are listed for easy reference.
- Teaching methods. This is an outline of the activities during the class, and the methods used to achieve the objectives.

## Activities

Many of the items listed under teaching methods refer to activities. These may be games, discussions, role plays, field trips, etc., as described in the next section. For each activity, an activity sheet has been included in Chapter 12. The activity sheets explain the purpose of the activity, how to run it, what equipment is needed, and suggested follow-up questions for discussion.

## Teaching techniques

There are two main ways that people learn. The first is through being told something by somebody else. As we all know, it is easy to forget something that you are told. The second is to discover something on your own, either by doing something new, or using things that you already know to understand something in a new way. When we discover new knowledge on our own, it is much easier to remember. This is partly because it is fun, and gives us a feeling of satisfaction.

These two types of learning are described as 'teacher-centred learning', where all the knowledge comes from the teacher, and 'student-centred learning' where the knowledge is either discovered by the students or comes from a new understanding of things they already know. In many societies, teacher-centred learning is the most common way that people are expected to learn. Listening to lectures, taking notes from a blackboard, and memorising lists of things is an approach that has been successfully used for years. However, there are two problems with it. Firstly, it is not fun, and because the facts that are being learnt are not connected to anything, it is easy to forget. The second problem is that things are not placed in context when they are taught. This means that it is harder for the students to use the facts in a real-world situation to solve problems.

Student-centred learning starts with the students' own experience of problems in the real world. Guided by the teacher, students are encouraged to come up with solutions to these problems, either using their own experience, sharing the experiences of their fellow students, or making new connections with knowledge they already have. Naturally, the teacher is required to provide new information. However, if the new information provided by the teacher is given when students actually want or need the information to solve a problem they are dealing with, then the new information immediately has a useful purpose and is placed in context. These new facts will not be easily forgotten, and can be used to solve other similar problems faced outside the classroom.

When either teacher or students are not used to using student-centred learning techniques, it can be quite difficult at the start. However, after trying for a while, both will realise that it makes teaching and learning more fun, and that the things being taught are useful. When running a training course, a quick check to see if you are using student-centred learning techniques is to listen for a while. If the students are doing most of the talking, then it's working the right way. If the teacher is doing most of the talking, something is wrong.

How, then, can a teacher encourage student-centred learning? The main technique is to use the experience that the students already have, and to present them with problems that they have already faced. To solve these problems they need to think, discuss with other students and discover new information. A skilful teacher is able to guide students so that they rarely need to be taught anything at all – most of the time, they discover things for themselves. It is surprising how often students already have a basic intuitive understanding of apparently complex concepts.

A range of different techniques and suggestions is presented below, to help trainers use student-centred learning effectively.

## Learning land marks

Effective learning has more to do with organising information than memorising new information. If a student understands the relationships between the different things they already know and the new things they learn, then they are able to use that information to help with everyday tasks. Trainers need to help students organise the

information - how does the thing that is being taught now relate to other things that I already know and have previously learnt? How will I be able to use this knowledge?

When travelling on a journey, land marks help people navigate and know where they are. Learning land marks are pointers for students that show where they have come from, where they are now, and where they are heading. If students always know exactly where they are, it is much easier to organise the information. If they get lost, and don't know where they are going, or how this information is to be used, then they don't know how this information relates to other things they know, and they don't know how to properly organise it. Unless the connection is made, the information may be wasted.

There are three good ways to provide students with learning land marks. The first is to give them a map of the lesson, so they can chart a course. At the beginning of every day, or every training session, give the students a brief outline of what is going to be covered. Make sure that it is clear how each topic is related the previous or the next topic, and why it is relevant.

The second technique is to regularly fix your position along the way. As each topic is dealt with, make sure the students know where they are up to. Introduce the topic, and better still, tick off the previous topics on the board. As each new concept is introduced, provide one or two examples of how this is relevant to the real world. The examples used throughout this book are there to help the readers understand how the topic being discussed is related to the real world.

The third way to provide learning land marks is to look back over the journey when it has ended. At the end of each lesson, run through all the topics covered, and highlight how they relate to each other.

## **Reinforcement and practice**

Most subjects, including active surveillance for livestock diseases, use knowledge that is built up layer by layer. It is possible to learn the next level only once the previous one is well understood. If new information is taught before the earlier information is properly learnt, then the foundations become unsteady, and students can become confused.

A good way to ensure that all the earlier information is well understood is to continually reinforce and practice it. Every time students are asked to remember something they have learnt, and to use it to solve a new problem, it makes it harder to forget. The teacher should therefore take every opportunity to include previous concepts in new exercises and problems, to help students practice them.

## **Warmers**

In student-centred learning, the students are expected to do most of the work, while the teacher guides them, providing new information when it is needed, and giving them direction. At the start of a lesson, students are often not prepared to take this active role. They are not yet thinking about the problems that need considering, and they may feel shy or inhibited about speaking out in front of the group.

Warmers are exercises that are designed to 'warm up' the students, to start them thinking about the problems and topics to be dealt with, and to make them comfortable speaking aloud and discussing things with other students. Warmers should be relatively short exercises, that involve a lot of student activity, and above all, are fun. It is best to use a warmer that deals with issues from the previous lesson, so they can practice what they learnt, while preparing them for the upcoming topics.

Any of the following activities can be used as a warmer, but games and competitions are often the best. Warmers can also be useful during village interviews, to help livestock owners relax and feel comfortable about speaking out loud (see page 107). A survey team which has had experience with warmers during their training will be much better able to use them during village interviews.

## Questions and answers

The simplest way to get students to actively participate in the lesson is to ask direct questions. Questions may be asked of the group as a whole, or directed to individuals. Targeting individuals forces them to participate, and avoids the problem of nobody being willing to speak first.

Questions can be used in two ways. Firstly, questions are good for introducing a new topic. A question is posed on how to deal with a problem, (for instance, how to select animals, how to collect blood from a pig, how to get participation from women during a village interview). This can then lead into a full discussion of the issue, maybe using some of the other techniques listed here.

The other way questions can be used is to check whether a topic has been properly understood. Using a new context or different example students are asked to use the new information to solve a problem or explain one aspect of the topic. If the student is unable to do this, or makes mistakes, another student is asked to comment or help. If several or all students show that they don't understand, then the topic hasn't been adequately taught, and you will have to think of a better way to explain or practice the concepts.

Using questions to check students' understanding is a quick and simple way to evaluate the effectiveness of your teaching. As a trainer, the only way to improve is to understand where there are weaknesses in your training, and think of new ways to solve these.

## Games or competitions

Much of the training involves serious or complex issues. To make the training more enjoyable, games or competitions can be used. These allow students to relax, have fun, but still practice the ideas they have been learning, or learn new ones through the game.

Competitions can be effective warmers, such as the knowledge quiz (Activity 22). This provides students with an opportunity to recall and practice information they have learnt, as well as gaining a feeling of pride. It also can be used by the teacher to assess the level of understanding of concepts.

As well as warmers, games can be used to introduce or practice new concepts. The sampling jigsaw game (Activity 7) is an example of this. Through the game, students have an opportunity to see for themselves the effect of different sampling strategies, and have fun at the same time.

## Group discussions

In group discussions, the class is divided up into a number of smaller groups. Each group is asked to discuss a topic, or consider a number of questions, and record their ideas on a piece of paper as they go. At the end of an allotted time, one member of the group presents the findings to the rest of the class. Each group should be given an opportunity to present their ideas.

Group discussions are an opportunity for students to explore and discuss the issues with each other, to relate their own experiences and share those of others.

They are a good way for students to discover how much they already know about a subject.

Groups may be made up of just two people, but more usually have between 4 and 6. Before the discussion, you should make it very clear what the aim of the discussion is and what the topic or questions under consideration are. During the discussion, the trainer should wander from group to group, monitoring the topic, and checking that they haven't strayed onto something else.

While each group is reporting their ideas, they should be recorded on the board for all to see. At the end, the trainer needs to summarise and organise the ideas, to give structure to the conclusion.

## Brainstorming

Brainstorming is a technique in which class is given a topic, and students are asked for the first idea that comes into their head. It is used to collect a lot of ideas quickly and to encourage participation. Brainstorming exercises may be used as warmers, or to introduce a new topic.

For a brainstorming session, the teacher presents a question or idea. Students are then asked to respond to this question or idea with their own ideas, using just one or two words, and responding quickly. The teacher invites each person in the group to respond, and writes down the answers as they go. Students are told that there is no such thing as a wrong answer in a brainstorming session, it is just to collect lots of ideas.

To be successful, brainstorming should be done very quickly and with some excitement. You should choose the order of students in an unpredictable way, and jump to the next student quickly. Don't let any discussions or argument start at this stage, just collect the ideas.

When the list has been made and there are no more ideas, the list can be used as a basis for the next activity or session, depending on the objectives of the lesson.

## Ranking

Ranking activities are used to set priorities or arrange things in order of importance. Ranking may be used during village interviews to identify priority livestock diseases, but it can also be a useful tool during training courses.

There are many ways to run a ranking activity, and these can be adjusted to the specific situation. Normally, the activity starts with the creation of a list, for instance a list of the common diseases affecting village livestock in the study area. Participants are then asked to identify which of the listed diseases are the most important, and which are less important. It must be made clear what "important" means. You may choose to define importance as 'most likely to cause death', or 'causing the greatest financial loss' or 'causing the most inconvenience', or most expensive to treat. Separate ranking exercises can be conducted for each of these different criteria, if desired.

Each participant scores each disease on a piece of paper in order of importance. The most important disease gets a score of 1, the second most important disease gets a score of 2 and so on. When all participants have finished, the scores of all participants for each disease are added up. The disease with the lowest score is the most important, through to the disease with the highest score which is the least important.

One important use of ranking is to help participants identify their own preconceptions and biases. If a disease ranking exercise such as the one described is

carried out, and the same exercise is used then with village livestock owners during a trial interview, the differences in the diseases and ranks between the participants and the livestock owners can be highlighted. These differences represent differences in attitude or experience of the veterinary services compared with the livestock owners.

## Role playing

Role playing involves asking participants to act out some scene or situation. Role playing allows participants to think about issues that are likely to be raised during field work, and to develop appropriate ways of dealing with them while still in a safe, non-threatening environment. It is also a good way to get people participating and break down inhibitions.

A role play is like a very short play, acted by the participants. Each of the players is given clear instructions as to who their character is, and what position or attitude they have. The players then act out the play, making up the dialog as they go along. Usually, a role play involves some sort of conflict or disagreement, that the actors need to resolve.

Activity 17 is an example of a role play. Participants use the play to explore a situation that is likely to arise during field work. Livestock owners are reluctant to let the survey team collect blood from their animals, and the survey team needs to explain why they should let them.

## Field trips

The purpose of much of the training is to prepare participants for survey field work. By far the best way to do this is to actually do the work. Field trips give the participants an opportunity to practice the skills they have learnt, and to experience for themselves the problems and limitations of working in the field.

Field trips can provide valuable information for survey planning as well, as they can act as small pilot surveys. The activities of the survey staff, and the responses of the livestock owners can be assessed, and problems identified and corrected. However, a training field trip is necessarily very different from a real village visit during a survey. Participants are generally less confident with their new knowledge, and, more importantly, there will usually be many more participants on the field trip than would participate in a normal village survey visit.

Very good organisation is therefore essential if the field visit is to be successful. The purpose of the trip and activities to be carried out should be carefully explained, and the roles and responsibilities of each of the participants must be clearly assigned.

While carrying out training activities in front of livestock owners, the trainer should be aware of the sensitivities of both livestock owners and participants. For instance, if inexperienced participants are practicing restraining or collecting blood from an animal, don't let too many participants use the same animal, or the animals belonging to a single livestock owner. Both animal and owner are likely to become stressed.

It is also important to try to maintain the participants' status while working in the village. If participants are seen by livestock owners to be unskilled or ignorant, then both the owners' and the participants' confidence will be undermined. Teaching during field trips should therefore concentrate heavily on encouraging students to demonstrate and practice their knowledge and skills. It should be done in a positive, supportive way, avoiding direct criticism.

After the field trip, there should always be a time set aside for discussion. A few questions should be used to stimulate the group to talk about their experience. In particular, it is important to identify what problems were encountered, and how they could be addressed or avoided. Make a point of identifying the good things as well as the problems. Activity 18 describes a trial village visit.

## **Practical activities**

In addition to field visits, training should include as many, real-life practical activities as possible. The idea is for participants to learn 'on-the-job', and to feel that the activities they are doing during the training are not just made-up exercises, but are actually contributing to the aim of the work.

One example is the tasks involved in survey planning. After the principles have been taught and practiced with various exercises, they can be used, during the training, to prepare for the real survey. For instance, random selection of first-stage units (villages) in a two-stage prevalence survey can be done by the group during training. The participants work together to obtain and check the sampling frame, and then use the software to calculate sample size and select the required number of villages.

A similar approach can be used after the survey with data analysis. While the principles of analysis can be taught, analysing the actual data collected during field work will give the exercise much more meaning. There is also the advantage of having a larger group of people entering and analysing data, making it faster, and easier to use duplicate entry checking systems.

## **Field work**

Although not formally part of the training course, learning continues during the actual field work of any survey. The field work should start as soon as possible after the finish of the training course, and be seen as a logical extension of it. Each survey team should be encouraged to hold a brief meeting at the end of each visit, to discuss what problems occurred, and how the procedures can be improved to address these problems.

The trainer should participate in the field work alongside the survey team as much as possible. At a minimum, the trainer should accompany the team on a number of visits, especially during the early part of the survey. This is to help continue the training and refine the skills of the survey teams, and also to identify problems, errors, or poor practices that may have slipped in to the work routine. If these are corrected at the start of the survey work, the quality of the survey will not be compromised.

# 11

## Lesson Plans

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Course 1 - Active surveillance, survey planning and sampling

Course 2 - Field techniques for livestock disease surveys

Course 3 - Computerised data management and analysis, and reporting

Chapter 12: Activity Sheets



The lesson plans in this chapter are provided as a resource for trainers. They are divided into three separate courses.

Course 1, "Active surveillance, survey planning and sampling" is designed for national staff, survey planners and coordinators, to prepare them for the task of coordinating surveys.

Course 2, "Field techniques for livestock disease surveys" is designed for field staff and the survey teams. This course provides training in all the practical data collection activities required. The design of these courses assumes that the participants of Course 1 will also participate in Course 2, and that the two courses will be followed, almost immediately, by the actual field work of a survey.

Course 3, "Computerised data management and analysis, and reporting", is designed for national staff, and coordinators. The aim is to run this course for the same participants as Course 1, soon after the completion of field work. The data collected can then be used as material for training.

After the completion of the training courses, and participating in the field work, trainees should be in a position to organise and conduct further livestock disease surveys as required.

The lesson plans therefore provide a structured syllabus for teaching the techniques described in this book. However, they are not appropriate to every situation, and not every training course needs to cover all the material. While some trainers may wish to use the lesson plans much as they are presented, they can also be used simply to provide a guide and stimulate ideas for the running of similar courses. In particular, trainers should structure the lessons and use activities according to their preferences, and the needs of the participants.

# Course 1 - Active surveillance, survey planning and sampling

## Participants

National staff, survey planners and coordinators

## Course Structure

Lesson 1: Introduction to animal health information and surveillance . . . . .	2
Lesson 2: Surveys and inference . . . . .	3
Lesson 3: Measures of disease . . . . .	5
Lesson 4: Sampling . . . . .	6
Lesson 5: Sampling in practice . . . . .	7
Lesson 6: Sampling without a sampling frame . . . . .	8
Lesson 7: Practical RGCS field work . . . . .	8
Lesson 8: Introduction to survey planning . . . . .	8
Lesson 9: Trial survey - rabies vaccination . . . . .	8
Lesson 10: Prevalence surveys . . . . .	8
Lesson 11: Incidence rate surveys . . . . .	9
Lesson 12: Surveys to demonstrate freedom from disease . . . . .	9

## **Lesson 1: Introduction to animal health information and surveillance**

**Duration:** 2 hours

**Location:** Classroom

### **Objectives**

- Discuss the use and importance of information on animal diseases
- Examine how information is collected in the current system (passive surveillance)
- Identify weaknesses in the collection of information
- Introduce the concept of active surveillance to address these weaknesses

### **Key points**

- Many veterinary staff don't realise the importance of disease information to their jobs. Try to emphasise the relevance of information and the need for good information for the jobs of the participants.
- Passive surveillance systems suffer from under reporting and bias
- Active surveillance can overcome these problems

### **Page references**

Animal Disease Surveillance (page 14)

### **Teaching methods**

- Introduce the course
- Group discussion on the need for animal disease information. How do the participants use information in their own jobs? Who else needs information, and about what?
- Develop flow chart of collection of disease information. Have one participant draw the steps in information flow, while the others suggest different sources and paths.
- Use direct questions to investigate possible weak spots in the flow of information. Highlight the problem of under-reporting of diseases.
- Use group discussions to list possible reasons why a case of disease may not appear in national-level records
- Explain the term passive surveillance
- Use questions to find how it could be done better, and introduce the idea of active surveillance

## Lesson 2: Measures of disease

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Understand prevalence and how it is measured
- Understand incidence rate and how it is measured
- Explore the relationship between incidence rate and prevalence
- Consider examples of when to use incidence rate and when to use prevalence
- Understand the difference between clinical prevalence and seroprevalence
- Be able to interpret sensitivity and specificity as measures of a test's performance
- Understand the difference between apparent prevalence and true prevalence

### Key points

- Prevalence is the number of cases of disease at one point in time
- Incidence rate is the number of new cases of disease over a period of time
- Diseases of long duration have a higher prevalence
- Seroprevalence is easier to measure than clinical prevalence
- Make sure students understand basic principles of immunity, antibodies, and serological tests
- Diagnostic tests usually make a small number of mistakes
- Sensitivity is the proportion of true positives that a test detects, specificity is the proportion of true negatives
- Sensitivity and specificity can be used to correct for the mistakes of a test, and calculate true prevalence

### Page references

Measures of disease (page 26)

Diagnostic tests (page 32)

### Teaching methods

- Visual aids, examples and direct questions to introduce prevalence and incidence rate
- Example calculations
- Group discussion on which measure to use for two hypothetical situations (e.g. evaluate losses caused by Swine Fever, and monitor effectiveness of movement control regulations which dictate that all animals passing checkpoints must have been vaccinated)
- Activity 2: Sensitivity and specificity
- Questions and example on apparent prevalence versus true prevalence

## Lesson 3: Surveys and inference

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Explain the principle of a survey
- Introduce the concepts of population, and sample
- Contrast surveys with complete counting of the population (censuses)
- Explain the process of inference
- Define bias and explain the need for representative samples
- Discuss estimation and precision
- Identify the role of sample size on survey accuracy

### Key points

- Surveys examine only a small sample of the population
- The sample is used to make inferences about the population
- Inference can be wrong, giving a biased result
- Representative samples ensure that inference is not wrong

### Page references

Disease Surveys (page 18)

### Teaching methods

- Activity 1: Classroom census and survey for average age
- Questions on population and sample
- Explain inference using visual aids, stressing that a survey estimate can be wrong
- Explain bias, accuracy and precision using visual aids
- Activity 3: Biased sampling survey.
- Activity 4: Sample size effect and surveys
- Group Discussion - how are samples selected now? Are the samples representative? Ask participants to list the different ways they have selected samples in previous work. Have them consider potential for bias.
- Discuss techniques, highlighting potential for bias

## Lesson 4: Sampling

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Understand the need for random sampling to reliably select a representative sample
- Be able to distinguish probability from non-probability sampling techniques
- Select random numbers using physical randomisation, random number tables, and a computer
- Introduce the concepts of 'probability proportional to size sampling' and 'stratified sampling'
- Understand the requirements of a good sampling frame.

### Key points

- Random sampling is the only way to reliably select a representative sample
- In simple random sampling, all elements have the same probability of being selected
- Computers can simplify the task of selecting a random sample
- Systematic sampling can sometimes be used to avoid the need for a sampling frame
- Sampling frames should include every member of the population, once only
- The sampling frame determines the level of inference

### Page references

The need for random sampling (page 38)  
Random sampling techniques (page 40)  
Sampling frames (page 49)

### Teaching methods

- Assess the level of understanding of basic probability with questions
- Introduce the concept of chance and probability
- Examples, using dice, cards, coins of a random outcome
- Explain how we don't know the outcome of a single trial, but in the long run, we can predict what will happen over many trials
- Activity 5: Random numbers. Predicting the outcome
- Demonstration of using a random number table and computer generated random numbers
- Activity 6: Classroom age survey using random sampling
- Discuss repercussions of an incomplete sampling frame or one with duplications
- Activity 7: Sampling Jigsaw game

## Lesson 5: Sampling in practice

**Duration:** 2 hours

**Location:** Classroom

### Objectives

- Consider the problems of sampling from large populations
- Understand the principles and advantages of two-stage sampling
- Practice the actual selection of herds or villages from a sampling frame using a computer
- Explain the meaning of replacement and without replacement sampling

### Key points

- Building a sampling frame for large populations is usually too expensive or not possible
- Two-stage sampling removes the need for a complete sampling frame, and makes field work easier
- Computers can be used to select a sample from a sampling frame.

### Page references

Sampling from a Sampling Frame (page 41)

Two-stage Sampling (page 64)

### Teaching methods

- Questions about problems with sampling from large populations.
- Explain the benefits of two-stage sampling
- Activity 8: Selecting sample villages for survey (if using SRS or PPS)

## Lesson 6: Sampling without a sampling frame

**Note: This lesson is necessary only for surveys using random geographic coordinate sampling.**

**Duration:** 2 hours

**Location:** Classroom

### Objectives

- Consider the problem of random sampling in the absence of a sampling frame
- Introduce the technique of random geographic coordinate sampling
- Use the computer to select random coordinates
- Note how remotely sensed data can be used to screen points
- Demonstrate the use of a GPS unit
- Introduce the field procedures for selecting villages

### Key points

- RGCS is difficult and time consuming
- Remotely sensed data (satellite images or aerial photographs) can be used to make it less difficult
- The GPS uses satellites to pinpoint the ground location
- Villages must be identified with a single unique point
- The selection radius can be chosen during field work to make the task simpler

### Page references

- Selecting random coordinates (page 65)
- Identifying selected villages (page 71)

### Teaching methods

- Group discussion - sampling without a sampling frame. Consider the problems of ensuring that every village has the same chance of selection, and what techniques could be used
- Introduce RGCS
- Demonstrate GPS
- Practice using the computer to select random points
- Explain field procedures and the use of forms, using visual aids

## Lesson 7: Practical RGCS field work

**Note: This lesson is necessary only for surveys using random geographic coordinate sampling.**

**Duration:** 1 day

**Location:** Field

### Objectives

- Practice the techniques of selecting villages using RGCS

### Key points

- Trial selection of villages, or selection of some of the genuine survey villages
- Requires good planning, GPS, adequate transport, data recording sheets and computer for selection of points

### Page references

Selecting random coordinates (page 65)

Identifying selected villages (page 71)

### Teaching methods

- Preparation for fieldwork, explain activities and responsibilities
- Select actual villages using RGCS. Requires transport
- Analysis of GPS data, selection of villages
- Discussion of problems and difficulties

## Lesson 8: Introduction to survey planning

**Duration:** 2 hours

**Location:** Classroom

### Objectives

- Consider the steps involved in running a survey
- Understand the process of framing the question, and determining how to answer it
- Understand the factors that influence sample size considerations
- Understand the value of pilot surveys
- Consider issues of analysis and reporting before the start of the survey
- Plan a trial survey

### Key points

- The survey question must be able to be answered using a measurable value
- Variance, precision and confidence are important factors in determining sample size

### Page references

Outline of survey procedures (page 143)

### Teaching methods

- Group discussion on what the major steps are for running a survey
- Organise and add any missed steps
- Group discussion on the effect of variance. Present an example of two populations, such as one class in a school, and all the people in a village. Consider a survey to estimate the average age in each of the populations. How many people would be needed from the school class, and how many from the village?
- Group discussion of things that need to be arranged before a survey
- Compare with checklist
- Explain the need for pilot survey. Examples of what can go wrong in a survey, that would be avoided by a pilot
- Questions to stimulate thought on the importance of reporting

## Lesson 9: Trial survey - rabies vaccination

**Duration:** 1 day

**Location:** Classroom and area of town/city in which training is taking place

### Objectives

- Plan, implement and analyse a real survey
- Develop an appropriate question and way to answer it
- Practice concepts of building a sampling frame
- Carry out random selection from the frame
- Practice interview skills and data collection
- Perform simple data analysis
- Practice oral reporting skills

### Key points

- This is the first real-world survey to be conducted during the course. Good organisation is important to maintain the confidence of the participants
- Any other appropriate disease or species may be used depending on the situation, although some changes in design may be necessary

### Page references

Outline of survey procedures (page 143)  
Sampling (Chapter 3)  
Prevalence (page 26)

### Teaching methods

- Explain the activities carefully beforehand
- Activity 9: Prevalence of urban rabies vaccination survey
- Presentation of results by different groups
- Discussion of problems encountered during the survey

## Lesson 10: Prevalence surveys

**Note:** This lesson is necessary only if prevalence surveys are planned.

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Understand the basic steps in carrying out a two-stage prevalence survey
- Be able to decide on the best survey design to use in a given situation
- Calculate sample size, and understand the factors that are necessary
- Decide on appropriate stratification variables
- Select herds or villages from a sampling frame

### Key points

- The design chosen depends on the sampling frame available
- Variance and prevalence estimates are required for the sample size calculation. These usually need to come from previous surveys.
- The computer can be used to calculate sample size
- Details of second stage sampling are discussed in the next training course

### Page references

Prevalence Surveys (Chapter 7)

### Teaching methods

- Questions: Revise the need for two-stage sampling in large populations
- Questions: Revise the concept of prevalence and when to use it
- Present examples of different situations, and ask questions about how to carry out the survey: sampling frame available versus no sampling frame
- Use the survey design selection flow chart (page 151) to explain how to choose the right design
- Use the computer to calculate sample size. List the parameters necessary for the calculation, and have the group decide on appropriate parameters. Discuss their choices and your recommendations.
- Group discussion or questions on what an appropriate stratification variable would be
- Highlight the problems of selecting animals in the village, using examples. The solution to these problems will be discussed during the second training course (Course 2, Lesson 7).

## Lesson 11: Incidence rate surveys

**Note:** This lesson is necessary only if incidence rate surveys are planned.

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Appreciate the problems of collecting incidence rate measures
- Appreciate difficulties of remembering events from many years ago
- Know the limitations on which diseases may be studied
- Understand the procedure for carrying out a retrospective disease outbreak survey
- Use a computer to calculate sample sizes for a disease outbreak survey
- Know the key pieces of data that must be collected
- Understand the concept of two-sample analysis
- Be able to identify appropriate data sources for two-sample analysis

### Key points

- Incidence rate can be measured at the animal-level or village/herd level. Village/herd level incidence rate is easier to measure, and often more relevant to disease control programs.
- Surveys of past events are reliable only if the events are easily remembered
- Sample size calculation for disease outbreak surveys depends on the difference in the average time since the last outbreak.
- Interview procedures for collecting the information will be covered in the next course (Course 2, Lesson 5)
- Two-sample analysis requires two different, independent sources of information on village or herd disease outbreaks.
- These can be used to estimate the total number of disease outbreaks

### Page references

Incidence Rate Surveys (Chapter 8)

### Teaching methods

- Revise the meaning of incidence rate and the difference between incidence rate and prevalence
- Use examples to distinguish between animal-level and village/herd-level incidence rate
- Describe the survey procedure for village disease outbreak surveys
- Use examples to explain how sample size is calculated
- Use a computer to practice calculating sample size
- Use visual aids and the example of fish in a lake to explain the principle of two-sample analysis
- Use questions and examples to determine what sort of data sources are independent

## Lesson 12: Surveys to demonstrate freedom from disease

**Note: This lesson is necessary only if surveys to demonstrate freedom from disease are planned. The concepts covered are more complex than in other lessons.**

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Understand the situations when it may be necessary to demonstrate freedom from disease
- Understand the problems of using sampling and imperfect tests to demonstrate freedom from disease
- Understand the concept of minimum expected (maximum acceptable) prevalence
- Understand Types I and II error and their importance in survey design
- Understand the steps in a single-stage survey
- Be able to calculate sample sizes using a computer
- Understand how to use two-stage sampling in large populations
- Understand the concept of clustering of disease
- Be able to calculate optimal sample sizes for two-stage surveys (advanced groups only).

### Key points

- It is not possible to prove that a population is free from disease, if using imperfect tests
- Surveys can demonstrate that there is a low probability that, if the disease exists, the prevalence is greater than a specified level
- Two-stage sampling can be used for large populations, and populations with disease clusters

### Page references

Surveys to Demonstrate Freedom from Disease (Chapter 9)

### Teaching methods

- Group discussion on freedom from disease and when it may be necessary to be able to demonstrate it
- Revise the concept of sensitivity and specificity
- Activity 10: Classroom survey to demonstrate freedom. Use to stress that proof is impossible, and there must be a maximum acceptable prevalence
- Use a computer to calculate sample size for single stage surveys
- Give examples of how disease clusters in a population
- Questions on how to survey a large population (two-stage sampling)
- For advanced groups, demonstrate optimal two-stage sample size calculation
- For advanced groups, discuss the effect of combining tests in series or parallel

## Course 2 - Field techniques for livestock disease surveys

### Participants

Field staff, survey teams (including national staff, survey planners and coordinators from Course 1). The first three lessons cover much of the same material as presented in the first lessons of Course 1. At the end of the first course, split the participants into three groups, and have each group prepare one lesson. The members of the group can then share the responsibility of running the lesson and presenting the material. Be sure that they have a few days to prepare.

### Course structure

Lesson 1: Introduction .....	10
Lesson 2: Surveys and inference .....	10
Lesson 3: Random sampling .....	10
Lesson 4: Village interviews .....	11
Lesson 5: Ranking and village outbreaks .....	13
Lesson 6: Trial village interview .....	14
Lesson 7: Selecting random animals .....	14
Lesson 8: Animal restraint and blood sampling .....	14
Lesson 9: Trial village visit (interview and specimen collection) .....	15
Lesson 10: Preparation for field activities .....	15

## Lesson 1: Introduction

**Note:** Parts of this lesson may be presented by the participants from Course 1.

**Duration:** 2 hours

**Location:** Classroom

### Objectives

- Discuss the use and importance of information on animal diseases
- Examine how information is collected in the current system (passive surveillance)
- Identify weaknesses in the collection of information
- Introduce the concept of active surveillance to address these weakness
- Appreciate the balance between data quality and ease of collection
- Be able to identify appropriate sources of data for different questions
- Understand the advantages of village interviews for rapid, reliable data collection

### Key points

- The first part of this lesson is summary of Lesson 1 in Course 1, explaining the need for information on animal diseases
- Passive surveillance systems suffer from under-reporting and bias
- Active surveillance can overcome these problems
- Village interviews of livestock owners can draw on the collective experience and memories of all livestock owners, and get good quality information in a short time

### Page references

Animal health information (page 14)  
Types of data and quality of data (page 84)

### Teaching methods

- Introduce the course
- If appropriate, invite participants from the first course to lead some of the sessions. Training is one of the best ways of learning. Monitor the performance closely, and correct any mistakes, being careful not to undermine the confidence of the presenter.
- Group discussion on the need for animal disease information. Use direct questions to investigate possible weak spots in the flow of information. Highlight the problem of under-reporting of diseases.
- Explain the term passive surveillance
- Use questions on how to improve surveillance; introduce active surveillance
- Group discussion on possible data sources for animal disease information. Nominate sources, and discuss aspects of reliability, and difficulty of collection.
- Use examples of different types of information and ask questions to identify the best data source to use

## Lesson 2: Surveys and inference

**Note:** Parts of this lesson may be presented by the participants from Course 1.

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Explain the principle of a survey
- Introduce the concepts of population and sample
- Contrast surveys with complete counting of the population (censuses)
- Explain the process of inference
- Define bias and explain the need for representative samples
- Discuss estimation and precision
- Identify the role of sample size on survey accuracy

### Key points

- Surveys examine only a small sample of the population
- The sample is used to make inferences about the population
- Inference can be wrong, giving a biased result
- Representative samples ensure that inference is not wrong

### Page references

Disease Surveys (page 18)

### Teaching methods

- If appropriate, invite participants from the first course to lead some of the sessions. Training is one of the best ways of learning. Monitor the performance closely, and correct any mistakes, being careful not to undermine the confidence of the presenter.
- Activity 1: Classroom census and survey for average age
- Questions on population and sample
- Explain inference using visual aids, stressing that a survey estimate can be wrong
- Explain bias, accuracy and precision using visual aids
- Activity 3: Biased sampling survey
- Activity 4: Sample size effect and surveys
- Group discussion - how are samples selected now? Are the samples representative?
- Discuss techniques, highlighting potential for bias

## Lesson 3: Random sampling

**Note:** Parts of this lesson may be presented by the participants from Course 1.

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Understand the need for random sampling to reliably select a representative sample
- Be able to distinguish probability from non-probability sampling techniques
- Select random numbers using physical randomisation, random number tables, and a computer
- Introduce the concepts of 'probability proportional to size sampling' and 'stratified sampling'
- Understand the requirements of a good sampling frame

### Key points

- Random sampling is the only way to reliably select a representative sample
- In simple random sampling, all elements have the same probability of being selected
- Computers can simplify the task of selecting a random sample
- Systematic sampling may be used to avoid the need for a sampling frame
- Sampling frames should include every member of the population, once only
- The sampling frame determines the level of inference

### Page references

The need for random sampling (page 38)

Random sampling techniques (page 40)

Sampling frames (page 49)

### Teaching methods

- If appropriate, invite participants from the first course to lead some of the sessions. Training is one of the best ways of learning. Monitor the performance closely, and correct any mistakes, being careful not to undermine the confidence of the presenter.
- Assess the level of understanding of basic probability with questions
- Introduce the concept of chance and probability
- Examples, using dice, cards, coins of a random outcome
- Explain how we don't know the outcome of a single trial, but in the long run, we can predict what will happen over many trials
- Activity 5: Random numbers. Predicting the outcome
- Demonstration of using a random number table and computer generated random numbers
- Activity 6: Classroom age survey using random sampling
- Discuss repercussions of an incomplete sampling frame or one with duplications

## Lesson 4: Village Interviews

**Duration:** 3 hours

**Location:** Classrooms

### Objectives

- Be able to organise a village interview of livestock owners
- Identify people with the right skills to lead an interview
- Be able to use techniques to get good information from village livestock owners
- Understand how to encourage all livestock owners to participate in the interview
- Know the appropriate order for conducting an interview
- Be aware of potential problems that may arise, and how to address these problems during the introduction
- Build an animal sampling frame during a village interview

### Key points

- The aim is to have all village livestock owners attend the village interview
- Good organisation is needed to ensure that villagers know when the meeting is and that they attend
- Obtaining good quality information requires skill and practice
- Every effort should be made to ensure that livestock owners are happy to participate in this survey and any future surveys
- The introduction to the interview can be used to avoid problems, by explaining the purpose of the survey and addressing owners' concerns
- Building a complete animal sampling frame requires careful and persistent questioning

### Page references

General guidelines (page 104)

Introduction (page 114)

Building a village sampling frame (page 115)

### Teaching methods

- Use questions to revise the advantages of using a village interview to collect information
- Group discussion - problems with interview data and possible ways to overcome these problems
- Group discussion - problems with cooperation, during this interview, and in the future, and ways to overcome
- Present the typical order of an interview
- Activity 11: Role play - introduction to village interview
- Use questions to revise the idea of a sampling frame
- Demonstrate data collection forms for building a sampling frame
- Activity 12: Build a mock sampling frame in the class

## Lesson 5: Ranking and village outbreaks

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Rank disease priorities or other information from the village
- Understand techniques for determining the date of a disease outbreak in the past
- Be able to use village histories and village calendars to help livestock owners determine the date of an outbreak
- Understand the need for censoring times when collecting village outbreak data, and how to collect them

### Key points

- Ranking can be used to determine the importance of different diseases
- When collecting information about outbreaks, make sure that the livestock owners clearly understand the disease being discussed
- Village histories can help identify the year of an outbreak
- Village calendars can help identify the month or season of an outbreak
- Censoring times must be collected for villages that have had no outbreaks

### Page references

- Ranking disease priorities (page 116)
- Collecting outbreak history information (page 118)

### Teaching methods

- Activity 13: Disease ranking
- Discussion on what criteria may be used to determine which diseases are important
- Demonstration of data recording sheets for disease ranking
- Example of asking for the usual months for different diseases
- Activity 14: Retrospective questions
- Examples of building village histories and calendars
- Demonstrate and explain the use of the data recording sheet for village outbreaks
- Discuss ways to determine censoring times for villages with no outbreaks

## **Lesson 6: Trial village interview**

**Duration:** Half day

**Location:** Village

### **Objectives**

- Gain experience in conducting a village interview
- Develop confidence addressing livestock owners
- Present an introduction which addresses owners' concerns
- Build a village livestock sampling frame
- Determine disease priorities in the village
- Determine the normal time of occurrence for important diseases
- Determine the date of the most recent outbreak of a particular disease

### **Key points**

- Careful organisation is important. The village must be notified beforehand, and the best time for a meeting decided. Unlike a normal interview, there will usually be a large number of participants involved. Each must clearly understand their role and responsibilities.

### **Page references**

Village Interviews (Chapter 5)

### **Teaching methods**

- Preparation for visit, assigning responsibilities
- Activity 15: Trial village interview
- Discussion of successes, problems, and recommended solutions

## Lesson 7: Selecting random animals

**Duration:** 2 hours

**Location:** Classroom

### Objectives

- Understand how to select random animals, using both a random number table, and a computer
- Be able to identify individual animals that have been selected from the sampling frame
- Be aware of the possible concerns of livestock owners, and address these concerns convincingly

### Key points

- Random selection of animals can take place during the village interview
- To ensure a representative sample, it is important to avoid replacing animals when not necessary, and to follow the selection procedure carefully

### Page references

Sampling animals within a village (page 54)

### Teaching methods

- Use the sampling frame that was made during the village visit
- Demonstration - select random animals using a random number table
- Demonstration - select animals using the computer
- Explain how to identify individual animals
- Discuss the problem of bias if the survey staff count the animals
- Activity 16: Selection of animals from a village sampling frame
- Activity 17: Role play of owners' concerns during animal selection

## **Lesson 8: Animal restraint and blood sampling**

**Duration:** Half day

**Location:** Classroom, field

### **Objectives**

- Practice restraint and blood collection from all relevant species
- Understand how to handle, process and transport specimens that have been collected

### **Key points**

- Animals are usually more nervous around strangers. Have owners capture animals first
- Restraint is easier if done quietly, and without scaring the animal
- Good restraint is important to avoid injury to the animal or the field staff
- When collecting blood, make sure that enough has been collected

### **Page references**

Animal Restraint (page 87)

Specimen Collection and processing (page 95)

### **Teaching methods**

- Discuss and demonstrate restraint and sample collection equipment in the classroom
- Have participants handle equipment, and practice its use
- Practice techniques with real animals. The animals should preferably be research animals or purchased for the purpose.
- Demonstrate the technique once first, then have each participant practice
- Ensure that animals are treated humanely, and demonstrate this through your own behaviour
- Demonstrate sample processing, and transport techniques, and have the participants practice for themselves

## **Lesson 9: Trial village visit (interview and specimen collection).**

**Duration:** 1 day

**Location:** Village

### **Objectives**

- Implement all activities of a field visit
- Improve interview skills through practice, and implement suggested changes arising out of previous interview
- Select animals from a sampling frame
- Invite owners to submit animals for sample collection
- Restrain animals, and collect blood samples
- Process specimens appropriately

### **Key points**

- The field trip should be as similar as possible to real survey field work

### **Page references**

Survey principles and specific techniques (all of Parts II and III)

### **Teaching methods**

- Preparation, assigning roles
- Activity 18: Trial village visit
- Discussion of successes, problems, and suggested improvements

## **Lesson 10: Preparation for field activities**

**Duration:** Half day or more

**Location:** Class and elsewhere

### **Objectives**

- Ensure that all practical preparations for either pilot study or real field work have been completed

### **Key points**

- Stress the need for good planning and organisation
- Completing and storing data recording sheets should get special attention

### **Page references**

Survey procedures (page 143)

### **Teaching methods**

- Group discussions to develop checklist of all activities and preparations that need to take place before survey
- Prepare list of tasks
- Assign tasks
- Complete preparations

# Course 3 - Computerised data management and analysis, and reporting

## Participants

National staff, and coordinators

## Course structure

Lesson 1: Introduction and review of field work .....	15
Lesson 2: Principles of analysis and measures of disease .....	16
Lesson 3: Introduction to computers .....	17
Lesson 4: Data processing procedures .....	17
Lesson 5: Simple data analysis, descriptive statistics .....	18
Lesson 6: Prevalence surveys data analysis .....	18
Lesson 7: Incidence rate surveys – retrospective disease outbreak surveys .	19
Lesson 8: Incidence rate surveys – analysis of two data sources .....	19
Lesson 9: Surveys to demonstrate freedom from disease .....	19
Lesson 10: Reporting .....	20

## Requirements

This course uses computers in almost every lesson. Participants must have access to computers, preferably with no more than two people to each computer.

## **Lesson 1: Introduction and review of field work**

**Duration:** 2 hours

**Location:** Classroom

### **Objectives**

- Review fieldwork activities
- Suggest improvements for future work

### **Key points**

- This lesson is an opportunity for the participants to use their knowledge and experience to improve future work

### **Page references**

None

### **Teaching methods**

- Introduce the course and explain the content
- Group discussion on the fieldwork - strengths and weaknesses, suggestions for improvement - Ask groups to address each aspect of the field work in turn: training, preparation, village selection and so on.
- Be sure to record all suggestions, and act on them for future work

## Lesson 2: Principles of analysis and introduction to computers

**Note:** This lesson is necessary only for participants with no experience of using computers

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Understand measures of central tendency (especially mean)
- Understand measures of spread (especially variance or standard deviation)
- Become familiar with computer hardware
- Recognise the purpose of different types of major software
- Understand the storage of data in a computer database table
- Compile all data collection forms from the survey

### Key points

- Analysis of data converts a large amount of hard to understand data into a few easy to understand numbers (measures), which can be used to understand the disease. Mean and standard deviation are two examples of measures
- Computers use a set of instructions (programs) to process data.
- Data are stored in tables, made up of records (rows) and fields (columns)

### Page references

Measures of disease (page 26)

Principles of data management and analysis (page 124)

### Teaching methods

- Review measures of disease
- Activity 19: Analysis of data on age of participants
- Practical demonstration of the parts of a computer and the purposes of the different components. Open a computer and identify components, peripherals and different types of disks
- Demonstrate different types of major software
- Discuss different data types
- Use questions to help participants identify the most appropriate way to store different types of data
- Use visual aids to explain data storage (databases, tables, fields, records)
- Collect all data collection forms from the survey and check for completeness

## Lesson 3: Data processing procedures

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Check data for completeness
- Perform any manual coding necessary
- Understand how to deal with missing data
- Create a table
- Enter data into a table
- Check data for errors
- Manipulate data

### Key points

- If participants are familiar with a database program, use that program. If they are not, use any program that is available, and the trainer is familiar with. Epi Info is the recommended choice, as it can be made freely available to all participants.
- Don't go into detail on how to check data for errors after data entry. The techniques will be taught in the next lesson, and are the same as those used for simple data analysis.

### Page references

Data Processing Procedures (page 127)

### Teaching methods

- Activity 20: Data checking and data entry
- Most of this lesson should be taken up with individual or pair work on computers, with brief breaks to explain the procedures required for different operations
- Training will be much easier if several tutors experienced in the use of computers are available to help answer participants' questions

## Lesson 4: Simple data analysis, descriptive statistics

**Duration:** 3 hours

**Location:** Classroom

### Objectives

- Calculate means, standard deviation and confidence intervals for village level data
- Calculate proportions and confidence intervals for village data
- Generate frequency tables and cross-tabulations
- Analyse subsets of the data
- Create graphs of the results of analysis

### Key points

- Simple descriptive statistics can be generated quickly using the computer
- Analysing and graphing data in different ways gives a more complete understanding of the data

### Page references

Epi Info Manual (recommended)

### Teaching methods

- Perform simple data analysis
- Most of this lesson should be taken up with individual or pair work on computers, with brief breaks to explain the procedures required for different operations
- Training will be much easier if several tutors experienced in the use of computers are available to help answer participants' questions

## Lesson 5: Prevalence surveys data analysis

**Duration:** 3 hours or more (depending on data entry time)

**Location:** Classroom

### Objectives

- Estimate the prevalence and calculate confidence interval for the prevalence estimate based on two-stage sampling data
- Understand how the data requirements and analysis differs according to the survey design
- Calculate true prevalence from apparent prevalence
- Compare the prevalence estimates from two different surveys

### Key points

- Different data are required depending on the different survey design
- Apparent prevalence may be quite different from true prevalence with poor tests or low prevalence levels

### Page references

Review of survey designs (Chapter 7)  
Prevalence Data Analysis (page 160)

### Teaching methods

- Enter, check and recode data from prevalence survey
- Demonstrate the use of the Prevalence program for analysis
- Use questions and examples to explain the interpretation of the program output
- Use the True Prevalence program to convert to true prevalence based on test performance

## **Lesson 6: Incidence rate surveys – retrospective disease outbreak surveys**

**Duration:** 2 hours

**Location:** Classroom

### **Objectives**

- Understand the difference between traditional incidence rate measures and the survival curve measure of incidence rate
- Generate a survival curve describing village outbreak experience
- Interpret the summary measures of the survival curve
- Compare the results of two surveys
- Understand the interpretation of the Hazard Ratio
- Understand the need to adjust for seasonal patterns
- Be familiar with options for more complex analysis

### **Key points**

- Incidence rate uses a single number to summarise disease occurrence. Village outbreak surveys use a curve (the survival curve) to summarise disease occurrence.
- Just as incidence rate can be compared and the difference measured, so can survival curves
- The hazard ratio measures the risk of disease outbreaks in one group compared with another
- If two surveys are conducted at different times of the year, and the disease shows a seasonal pattern, you need to adjust for this to avoid bias

### **Page references**

Review of survey design (Chapter 8)  
Data Management (page 175)  
Data Analysis (page 177)

### **Teaching methods**

- Enter, check and recode data from village outbreak survey
- Analyse data using Survival program.
- Analyse different data sets to explain the need to adjust for seasonal patterns
- Use questions to clarify the interpretation of survival curves

## **Lesson 7: Incidence rate surveys – analysis of two data sources**

**Duration:** 2 hours

**Location:** Classroom

### **Objectives**

- Match outbreaks from two sources
- Analyse data from two data sources to estimate the total number of disease outbreaks
- Use this estimate to calculate the incidence rate

### **Key points**

- Clear rules have to be established for matching outbreaks between two sources
- If the number of outbreaks appearing in both sources is small, the estimate will have very wide confidence intervals
- Incidence rate requires a knowledge of the size of the total population (the total number of villages in the study area)

### **Page references**

Review of survey design (page 183)  
Data analysis (page 187)

### **Teaching methods**

- In small groups, match outbreaks between the two sources and calculate totals
- Use the program Capture Recapture for analysis of the results
- Calculate incidence rate and the confidence interval

## **Lesson 8: Surveys to demonstrate freedom from disease**

**Duration:** 2 hours

**Location:** Classroom

### **Objectives**

- Analyse data to calculate the probability that an area is free from disease

### **Key points**

- Analysis requires a knowledge of test performance (sensitivity and specificity), as well as maximum acceptable prevalence and Type I and II errors.

### **Page references**

Herd or village survey data analysis (page 198)

Large area survey data analysis (page 204)

### **Teaching methods**

- Enter, check and recode survey data
- Use the FreeCalc program to analyse the data
- Use questions to clarify the interpretation of the results, and the meaning of the null and alternative hypothesis.

## Lesson 9: Reporting

**Duration:** 1 hour, plus homework

**Location:** Classroom

### Objectives

- Understand the need for reporting at different levels
- Consider the best way to communicate the results at these different levels
- Understand techniques for clear communication of results
- Prepare reports of the survey results

### Key points

- Survey results should be reported back to everybody who has participated and everybody who may need the results
- Reports should be simple, clear, and easy to understand, and targeted at the user
- Written reports may not be appropriate for all users

### Page references

Survey procedures (page 143)

### Teaching methods

- Group discussion of who might need the information from the survey. Highlight why different people might need the results (including livestock owners, and the international community)
- Group discussion of how to best communicate with the different users of the information
- Activity 21: Report preparation

# 12

## Activity Sheets

### Contents

Chapter 1: Introduction

#### **Part I: Background to Disease Surveys**

Chapter 2: General Principles of Animal Disease Surveillance

Chapter 3: Sampling

Chapter 4: Principles of Data and Specimen Collection

Chapter 5: Village Interviews

Chapter 6: Computerised Data Management and Analysis

#### **Part II: Survey Design and Analysis**

Chapter 7: Prevalence Surveys

Chapter 8: Incidence Rate Surveys

Chapter 9: Surveys to Demonstrate Freedom from Disease

#### **Part III: Notes for Trainers**

Chapter 10: Guide for Trainers

Chapter 11: Lesson Plans

Chapter 12: Activity Sheets



This chapter contains a set of activity sheets, which provide a guide to running different types of activities during training courses. The activities are directed at the specific training courses described in the last chapter, but can be used modified or unmodified for different types of training courses.

Each activity is followed by a number of questions that may be used for discussion, to focus the participants' attention on what they have been doing.

## Activity 1: Introductory classroom census and survey

<b>Location:</b> Classroom	<b>Duration:</b> 20 minutes
<b>Reference:</b> Course 1, Lesson 3 Course 2, Lesson 2	<b>Software:</b> None
<b>Objectives</b> Differentiate between a complete count (census) and a sample (survey)	<b>Concepts practiced</b> Sampling Estimation Inference
<b>Equipment and materials</b> Board for recording ages	

### Description of activity

Explain that the aim is to determine the average age of all the participants in the room, and that this will be done in two ways. First, ask a sample of a small number of people. Record their age in a column on the board. When finished, have the participants calculate the average of the ages.

Next ask everybody in the class in turn, and write their ages on the board. Calculate the average and compare to the first result.

### Questions for discussion

Which way took longer?

If there were 3000 people in the room, which way would be best?

Was the answer from the survey right?

Was the answer from the census right?

## Activity 2: Sensitivity and specificity game

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> Course 1, Lesson 2	<b>Software:</b> None
<b>Objectives</b> Introduce an understanding of sensitivity and specificity	<b>Concepts practiced</b> Sensitivity and specificity Probability
<b>Equipment and materials</b> Board for recording responses	

### Description of activity

The aim is to demonstrate how participants intuitively use the concepts of sensitivity and specificity all the time. It is demonstrated using a non-diagnostic example.

Invite one participant to the front of the room. Ask them to make a decision about some unknown characteristic of each of the other members of the group. For example, to judge whether they originally come from the capital city, they belong to some ethnic or cultural group, they come from some broad region (the north or the south). Whatever characteristic is chosen should be culturally appropriate (not cause discomfort), and should have several people in the room representing both options (some from the north, and some from the south). It should also be some characteristic that is not immediately possible to be sure of, but the observer will get it right for most people most of the time.

Draw a two-by-two table on the board, and record the responses. For each person, the participant must first decide (e.g. say "North"), and then the person in question responds if this is right or wrong.

At the end of the exercise, ask how well the group thought they went. Ask if the person was better at picking people from the north than people from the south.

Use the figures to calculate sensitivity and specificity, and show how these can be interpreted. If possible, repeat the activity using a different person to decide, and the same characteristic. This second person should not have heard the answers from the first time. Compare the performance of the two people.

### Questions for discussion

How does this relate to diagnostic tests?

How could we improve the proportion of correct decisions?

### Activity 3: Biased sampling

<b>Location:</b> Classroom	<b>Duration:</b> 20 minutes
<b>Reference:</b> Course 1, Lesson 3 Course 2, Lesson 2	<b>Software:</b> None
<b>Objectives</b> Understand how non-representative samples produce biased estimates	<b>Concepts practiced</b> Sampling Populations Bias
<b>Equipment and materials</b> Board for recording responses	

#### Description of activity

Conduct a survey of the participants to calculate their average age from a small sample. Before selecting the sample, chose a group that will be clearly biased. For example, if younger participants mostly speak English, and the older participants don't, explain how this survey is going to be conducted in English. Select several people and ask their age in English, only recording those who can respond in English. Another biased sample is to only ask the people with the higher positions or ranks in the group, as a sign of respect or in recognition of their greater experience. Other approaches can be used to select a biased group, with ages either greater or lower than the overall average.

Select the group, record the ages, and calculate the average. Compare this average with the result of the census used in Activity 1.

#### Questions for discussion

Why are the results different?

If selecting animals, how could a similar thing happen?

### Activity 4: Survey sample size

<b>Location:</b> Classroom	<b>Duration:</b> 20 minutes
<b>Reference:</b> Course 1, Lesson 3 Course 2, Lesson 2	<b>Software:</b> None
<b>Objectives</b> Recognise the effect of different sample sizes on the reliability of a survey	<b>Concepts practiced</b> Surveys, sampling Sample size Variance
<b>Equipment and materials</b> Board for recording results	

#### Description of activity

Conduct two classroom surveys, to calculate the average age of participants. In the first, use a small sample size, say 4 people. In the second, use a large sample size, almost all the people in the room.

Record the results, and calculate the average age from both. Contrast the results with the true average from a census (Activity 1).

If desired, calculate confidence intervals around the estimates.

#### Questions for discussion

Which survey produced the most accurate result?

Which was fastest and easiest?

Was either survey correct?

Do the confidence intervals contain the real average?

## Activity 5: Random numbers

<b>Location:</b> Classroom	<b>Duration:</b> 40 minutes
<b>Reference:</b> Course 1, Lesson 4 Course 2, Lesson 3	<b>Software:</b> None
<b>Objectives</b> Become familiar with physical randomisation procedures Introduce probability and the prediction of random outcomes	<b>Concepts practiced</b> Random numbers
<b>Equipment and materials</b> Pack of playing cards, dice, coin	

### Description of activity

First demonstrate the concept of a single random outcome, in several ways. Shuffle the cards, and ask one participant to select one at random. Ask another to say if it is black or red. Have another roll the dice. Ask another to say what the result is without looking. Have another flip a coin while another predicts if it will be heads or tails.

Stress that the outcome of all of these things is random, and that the individual result cannot be predicted.

Next demonstrate how the average of many outcomes can be approximately predicted.

Split the group into three. In the first group, have one person flip the coin 100 times while the rest record the results. In the second, have one person shuffle the cards, and another draw one at random, and record if it is black or white. Replace the card and repeat 50 times, recording the results. In the third group, roll one die 120 times, recording the results in a table.

Explain the activities first, and then, before doing them, ask each group to predict what the outcome will be (how many heads, how many tails etc.). Record the predictions on the board. Then conduct the trials, and record the final results on the board.

### Questions for discussion

Why can't you predict whether a single flip will be heads or tails?

With many trials how well were you able to predict what the average outcome would be?

Was it exactly as you predicted, or just close?

How could this be used for survey sampling?

## Activity 6: Random sampling using a random number table and computer

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> Course 1, Lesson 4 Course 2, Lesson 3	<b>Software:</b> Random Village, or Random number generator (e.g. Epi Info)
<b>Objectives</b> Select random samples using random numbers	<b>Concepts practiced</b> Sampling Random numbers
<b>Equipment and materials</b> Board Random number table Computer	

### Description of activity

Conduct a classroom survey to determine the average age of participants. Have one participant build a sampling frame on the board, by asking each person's name, and writing a number next to it from 1 to the total.

Divide the group in two. Have one half use a random number table to select a random sample (with a reasonably large sample size), and the other half use the computer to generate random numbers for the sample.

Conduct two surveys, using the sets of random numbers, and calculate the average ages.

### Questions for discussion

Were the results the same?

Were the results correct?

Are they closer to the real value than using other techniques?

Are they representative?

## Activity 7: Sampling jigsaw game

<b>Location:</b> Classroom	<b>Duration:</b> 1 hour
<b>Reference:</b> Course 1, Lesson 4	<b>Software:</b> Random Village
<b>Objectives</b> Implement 4 different sampling schemes Examine how non-random sampling causes problems with inference	<b>Concepts practiced</b> Convenience sampling Haphazard sampling Random sampling Systematic sampling Inference to the population
<b>Equipment and materials</b> Jigsaw puzzle, with 40 to 80 pieces, showing a scene with lots of variation Each piece should be numbered on the back sequentially from 1 to the total Four sets of pen and paper for four groups, <i>or</i> , overhead projector, and four sets of overhead transparencies and pens for four groups.	

### Description of activity

The group is divided into four groups. It is explained that we are going to use four different sampling techniques to select a sample from a population, and that the sample will be used to estimate what the population is like.

Use the Sampling Techniques figure to introduce or revise convenience, haphazard, systematic and random sampling.

In a separate part of the room, assemble the jigsaw puzzle, and turn it over so only the backs of the pieces with the numbers are showing. The population is all the pieces. The sampling frame is a list of numbers from 1 to the total population.

As a class, have each group select the pieces to sample, using four different sampling techniques:

The same number of pieces is chosen by each group. For a puzzle of about 50 pieces, each group chooses 5 pieces. For larger puzzles, select more pieces.

Group 1 uses convenience sampling, and selects pieces 1,2,3,4 and 5 (all from the top corner of the puzzle).

Group 2 uses haphazard sampling. Ask each member of the group to pick a number between 1 and the total. Use the first 5 numbers chosen.

Group 3 uses systematic random sampling. If there are 40 pieces, the sampling interval is  $40/5 = 8$ . Have the group use either a random number table or the Random Village program, to pick one number at random between 1 and 8, as the starting number. Then pick every eighth number.

Group 4 uses simple random sampling. Use a random number table or the Random Village program to select 5 numbers (without replacement) between 1 and the total.

One by one, have each group come to the puzzle, and examine the 5 selected pieces. Do not let the other groups see the front of the pieces selected. When the pieces have been examined, have each group draw a picture of what they think the jigsaw puzzle is showing. This is an example of inference, guessing what the population is like, based on a sample. Give each group 5 minutes to draw their picture. They may want to write on the picture to explain what certain things are.

When all groups are finished, show each of the pictures (preferably on an overhead projector), and note which sampling technique was used. Then show the complete jigsaw puzzle, so the class knows the true state of the population. Finally, have the class vote on which group's picture was most like the real population.

**Questions for discussion**

Talk about how well each sampling technique was able to provide a representative sample.

Usually, systematic sampling and random sampling provide a good picture. Haphazard may sometimes give a good picture, but convenience never does. If haphazard is better than simple random, discuss how simple random sampling does not *always* provide a representative sample, but that it is the only way to give a representative sample most of the time.

## Activity 8: Selecting actual villages from a disk based sampling frame

<b>Location:</b> Classroom	<b>Duration:</b> 1 hour
<b>Reference:</b> Course 1, Lesson 5	<b>Software:</b> Random Village
<b>Objectives</b> Select a sample of herds or villages from computerised sampling frame	<b>Concepts practiced</b> Use of the computer for random sampling
<b>Equipment and materials</b> Computer Village or herd sampling frame (with or without population figures) in either Paradox, or dBASE format	

### Description of activity

This is part of the preparation for the real survey. Discuss the source of the sampling frame, what information is included (population figures or not), how well herds or villages are identified, and how complete the list is likely to be.

Ask the group to consider options for stratification. The sample size should have already been calculated.

Demonstrate the use of the Random Village program, using a dummy data set.

Have several members of the group use the program to select the sample. Save and print the results, and have staff with local experience examine the list for any villages or herds that are impossible to survey. Use the program to replace these with new villages/herds.

### Questions for discussion

What is the effect on replacing inaccessible herds? How do the results of the survey suffer?

## Activity 9: Local survey - dog rabies vaccination

<b>Location:</b> Urban area	<b>Duration:</b> 1 day
<b>Reference:</b> Course 1, Lesson 9	<b>Software:</b> Random Village or Epi Info random number generator
<b>Objectives</b> Implement field survey	<b>Concepts practiced</b> Sampling and sampling frames Data collection and analysis
<b>Equipment and materials</b> Transport to survey site Data recording sheets	

### Description of activity

The aim of this activity is to implement all the newly learned skills in the form of a survey to estimate the proportion of dogs in the local urban area that have been vaccinated against rabies within the preceding year. Other species and diseases may be used if appropriate.

In the classroom, explain the aim of the survey. Have participants frame the question (e.g. "Is the local rabies vaccination campaign working?") and refine it to a measurable quantity ("What proportion have been vaccinated in the last year?"). Define the target population (all dogs in the local urban area), and the study population (probably a much smaller area that can be surveyed in a few hours).

Discuss the options for sampling frames and sampling strategies. Assuming that no list of dogs is available (use it if it is), a household frame can be used instead. If no household sampling frame exists, then one must be generated.

Select a small area nearby, consisting of one or two streets, or a small number of blocks, with a total of several hundred houses. Draw a sketch map of the area. Divide the group into pairs, assign each pair to a particular part of the study area. Travel to the area, and have each pair prepare a map of their section, showing all households (preferably identified by house number).

Return to the classroom, and compile these maps into a map of the overall study area. Number each household consecutively starting at 1.

Have the group select random numbers using either a random number table or computer, to select a sample of households. When selecting, stratify by section of the study area. The sample size should be as large as possible. The group will divide into pairs to do the survey work, and each pair should have about 15 to 20 households to survey. If there are 10 participants, there are 5 groups, and the sample size can be about 100.

Prepare data recording sheets. The questions should be kept to a minimum, and include "Do you have any dogs?", "If so, how many?", "For each dog, was it vaccinated against rabies during the last year?". Discuss the brief questionnaire and the need for a short explanation to householders of what the survey is about.

Return to the study area, split into groups and collect the data. When finished, return to the classroom for data analysis. Tally the total number of dogs and total vaccinated on the board, and calculate the proportion. Use Epi Info to calculate a

confidence interval for the proportion. Have participants prepare a half-page report of the survey, its findings and conclusions, suitable for submission to the local authorities.

**Questions for discussion**

Is the rabies vaccination campaign working?

Can we use inference to estimate the situation in the whole town/city?

What population do the results relate to?

What problems did you have?

How could the survey be improved?

## Activity 10: Survey to demonstrate freedom from disease

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> Course 1, Lesson 12	<b>Software:</b> FreeCalc
<b>Objectives</b> Understand the difficulty of proving freedom from disease with imperfect tests	<b>Concepts practiced</b> Sensitivity and specificity Freedom from disease
<b>Equipment and materials</b> Computer	

### Description of activity

Begin this activity with a revision of sensitivity and specificity. Then, using the same or similar characteristic as used in Activity 2 (Sensitivity and specificity), have one member of the group make a decision about a sample of the others. Ideally, the characteristic should be rare, but quite possible. One possibility is whether the person was born abroad or not.

The assessment of the one person is recorded on the board, but the true status is not yet revealed. When the decisions have been made, the class is asked whether the results indicate that there is nobody who was born abroad. Explain how a sample is unable to prove this, because not everybody has been checked.

Repeat the survey using everybody in the room, and record the decision.

You may extend this exercise to analyse the results with the FreeCalc program.

### Questions for discussion

How confident can you be of the final conclusion?

How could you improve your confidence?

## Activity 11: Introductions for village interviews

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> Course 2, Lesson 4	<b>Software:</b> None
<b>Objectives</b> Practice explaining the purpose of the interview, and address problems which are likely to be of concern	<b>Concepts practiced</b> Addressing groups Awareness of the concerns of others Avoiding problems
<b>Equipment and materials</b> None	

### Description of activity

Select one person to act as the leader of the interview. The rest of the group takes the role of the village livestock owners. Ask the leader of the interview to spend 5 minutes preparing the main points to be covered during the introduction to the interview – for example, why we are here, what we intend to do, etc. While they are preparing, ask the rest of the group to imagine that they are livestock owners. Have them think of various objections to participating in the interview or giving information. Ask them to try to make things difficult for the leader.

Start the role play, and ask the leader of the interview to introduce the survey, and try to explain things so as to avoid as many problems as possible. When they are finished, the livestock owners should ask questions or raise concerns about those issues which the leader has not explained.

Repeat the exercise with a different leader (perhaps choose the most vocal of the livestock owners to play the role).

### Questions for discussion

How well did the leaders explain the survey?

Were there any major points that were missed?

Are there any potential livestock owner concerns that can't be addressed during this introduction?

## Activity 12: Building a livestock sampling frame

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> Course 2, Lesson 4	<b>Software:</b> None
<b>Objectives</b> Develop a sampling frame by asking livestock owners about their livestock	<b>Concepts practiced</b> Interview skills
<b>Equipment and materials</b> Data record sheets for the sampling frame	

### Description of activity

Select one member of the group to be the leader of the interview. Have that person leave the room, and ask the rest of the group to play the role of livestock owners, and decide how many animals they each have. Then ask different livestock owners to adopt different personalities that may be difficult during an interview. Ask one person to act as if they know it all, and try to answer for others, or correct their mistakes. Ask another pair to be bored and want to chat all the time. Ask some to pretend to be deaf, some to be unsure of what is wanted, and some to be suspicious. Ask some to try to give the wrong information, without lying.

Ask the leader to return, and collect livestock information.

### Questions for discussion

How well did the leader do the job?

Is all the information correct?

Could they have done it any better?

## Activity 13: Disease ranking

<b>Location:</b> Classroom	<b>Duration:</b> 40 minutes
<b>Reference:</b> Course 2, Lesson 5	<b>Software:</b> None
<b>Objectives</b> Rank diseases in order of importance	<b>Concepts practiced</b> Ranking Interview skills
<b>Equipment and materials</b> Board Paper and pens	

### Description of activity

Divide into several small groups, and ask each group to list on paper all the important diseases that occur in the local area, in the species of interest. When finished, record all these diseases on the board. Next have the groups discuss how the importance of these different diseases may be assessed, and to write a list of 3 or 4 different criteria for ranking.

Write the different criteria on the board, and agree on 4 criteria to use to rank the diseases. Ask each person to rank the diseases according to each of the criteria, scoring the most important disease with 1, down to the least important.

Add up everybody's scores for each disease, and each criteria, and assign overall ranks.

### Questions for discussion

Will livestock owners rank the diseases in the same way?

Is a disease with a rank of 10 twice as important as a disease with a rank of 20?  
(No)

How could ranking be done in a more quantifiable manner?

## Activity 14: Retrospective questions

Compare the results from the different groups, and discuss how they remembered.

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> Course 2, Lesson 5	<b>Software:</b> None
<b>Objectives</b> Emphasise the difficulty in remembering past events, and understand how it can be made easier	<b>Concepts practiced</b> Group memory Dating landmarks
<b>Equipment and materials</b> Paper to record responses	

### Description of activity

Break into small groups. Identify an event that occurred some years ago that every participant knows about. Examples include elections, natural disasters, major news items, etc. Ask each group to try to remember the month and the year of the event.

Record the different responses from each group.

### Questions for discussion

Are the answers the same? If not, what sort of errors were made (wrong year, wrong month)?

How did participants remember?

How can remembering be made easier?

## Activity 15: Village interview

<b>Location:</b> Village	<b>Duration:</b> Half day
<b>Reference:</b> Course 2, Lesson 6	<b>Software:</b> None
<b>Objectives</b> Practice a village interview in the way it will be done during the real survey	<b>Concepts practiced</b> Public address Explanations Ensuring cooperation Encouraging participation
<b>Equipment and materials</b> Transport Data recording sheets, pens, paper	

### Description of activity

Make sure that the village interview has been organised beforehand, and that the livestock owners know when it is. Prepare the group well, so they know their responsibilities and roles.

Try to use a village with a large number of livestock owners. Divide the village into several smaller groups, so that as many as possible of the participants have the chance to lead the interview. There should be two or three participants running each interview, with a minimum of 5 or 6 livestock owners. The exercise is easier if there are a number of tutors available, experienced in village interviews to supervise each group.

Have the group run through the complete interview, as it will be run during the survey. For example, the following sections may be included:

- Explain the purpose of the interview
- Build a livestock sampling frame
- Rank disease problems
- Ask about the usual dates of disease problems
- Ask about disease outbreak history
- Invite questions on disease problems and offer advice

### Questions for discussion

What information did each group collect?

Did each group get the same answers to the same questions?

What problems were encountered? How could these be addressed?

## Activity 16: Random selection of animals

<b>Location:</b> Classroom	<b>Duration:</b> 40 minutes
<b>Reference:</b> Course 2, Lesson 7	<b>Software:</b> Random Animal
<b>Objectives</b> Use a sampling frame to practice selecting individual animals	<b>Concepts practiced</b> Random selection
<b>Equipment and materials</b> Sampling frames from village visit (1 copy for each participant) Random number table Computer Pen and paper for results	

### Description of activity

Copy the sampling frame generated during a village visit (e.g. Activity ?). Give a copy to each participant, with a random number table. Have each participant select a group of 10 animals at random. While they are doing this, have each student in turn use a computer and the Random Animal program to enter the sampling frame and select 10 animals. When the samples have been selected, use a role play to practice identifying the individual animals. Have one participant play the role of the survey team leader, another the livestock owner, and several others play the role of the animals. Select several animals, then move to a new owner, and new animals.

### Questions for discussion

Which technique was simplest for selecting animals, the random number table or the computer?

Which is most appropriate for use in the village?

How well did the team leader select the animals? Are there any suggested improvements?

## Activity 17: Specimen collection role play

<b>Location:</b> Classroom	<b>Duration:</b> 30 mins
<b>Reference:</b> Course 2, Lesson 7	<b>Software:</b> None
<b>Objectives</b> To understand livestock owners' concerns about specimen collection To practice addressing these concerns and encouraging cooperation	<b>Concepts practiced</b> Communicating with livestock owners
<b>Equipment and materials</b> Role Cards - one for each actor explaining their attitude	

### Description of activity

Four of the group are selected to participate in the role play. One person plays the role of the member of the survey team, and three play village livestock owners. The scene is explained. The member of the survey team has come to collect blood. They visit the livestock owners, and want to collect blood from one animal belonging to each owner.

Each of the actors is then given a Role Card to explain what position they are to take during the play.

The member of the survey team needs to collect blood from the selected animals. They are not allowed to change the animals for another, and must get blood.

The first livestock owner is happy for them to try to collect blood, but doesn't believe that they will be able to hold their buffalo still.

The second livestock owner doesn't want blood to be collected from the selected animal, because she is pregnant. They have seen vaccinated animals abort after being vaccinated.

The third owner doesn't want their animals to be used because they are very suspicious of what the information will be used for. They are afraid that the team is trying to prove that they are not looking after their animals properly.

Conduct the play, telling the actors that it should last no more than 10 or 15 minutes. Each of the participants should stick to their role, and try to argue for what they believe as strongly as possible.

At the end, the play can be repeated again, with different actors, or with the same actors in different roles.

### Questions for discussion

How well did the survey team member perform?

Are there any important points that they missed in their explanation?

What should you do if the livestock owners completely refuse to cooperate?

If you were the livestock owner, would you be happy to be involved in the survey if it was conducted again next year?

How can these problems be avoided?

## Activity 18: Village interview and specimen collection

<b>Location:</b> Village	<b>Duration:</b> 1 day
<b>Reference:</b> Course 2, Lesson 9	<b>Software:</b> Random Animal
<b>Objectives</b> Implement a complete village visit	<b>Concepts practiced</b> Interviews Restraint Blood collection
<b>Equipment and materials</b> Transport Data recording sheets Restraint and specimen collection equipment	

### Description of activity

Organise the visit carefully beforehand. Make sure everybody knows their role and responsibilities.

Divide into several groups and conduct the village interviews as described in Activity ?.

During the interview have one or two people select animals from the sampling frame.

Collect specimens from the selected animals. Initially, have all participants as one group while identifying and collecting specimens from the first few animals. Then, if tutors are available to supervise, split into smaller groups so that each person has an opportunity to practice every role in blood sampling - capturing, restraining, etc.

Process the specimens appropriately.

### Questions for discussion

What problems arose?

How could they be addressed?

## Activity 19: Analysis of data

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> Course 3 Lesson 2	<b>Software:</b> Epi Info or other database/statistical software
<b>Objectives</b> Understand basic principles of data entry and analysis	<b>Concepts practiced</b> Summarising data
<b>Equipment and materials</b> Computers (preferably 1 each or 1 between 2 participants)	

### Description of activity

Collect data on the age of each participant, and record it on the board. Using a table that has already been created, have each participant enter the data. Use EpiInfo Analysis program to calculate the

- mean,
- minimum,
- maximum,
- variance,
- standard deviation, and
- confidence interval around the mean.

### Questions for discussion

How is each of these values interpreted?  
What population does the data relate to?

## Activity 20: Data management

<b>Location:</b> Classroom	<b>Duration:</b> 3 hours
<b>Reference:</b> Course 3 Lesson 3	<b>Software:</b> Epi Info
<b>Objectives</b> Become familiar with data management procedures	<b>Concepts practiced</b> Table creation Data entry Data checking
<b>Equipment and materials</b> Computers	

### Description of activity

Using the raw data collected during village surveys, set the task of calculating basic village-level descriptive statistics on livestock population, proportion of villages suffering outbreaks, and the most important livestock diseases.

Lead participants through each of the operations step by step, then let them do it themselves, and explore the procedures.

Check data, create appropriate tables and data entry forms, set up checks during data entry, check data after data entry, recode data, export data, analyse data.

### Questions for discussion

How should the results be interpreted? What do the numbers say about the disease situation?

## Activity 21: Report writing

<b>Location:</b> Classroom	<b>Duration:</b> 2 hours
<b>Reference:</b> Course 3 Lesson 9	<b>Software:</b> None
<b>Objectives</b> Practice writing appropriate reports	<b>Concepts practiced</b> Reporting
<b>Equipment and materials</b> Paper, pens, possible word processor	

### Description of activity

After discussing the levels of reporting and appropriate types of report presentation for different audiences, assign several participants to work together to produce a short report for each of identified groups. Give them 2 hours to work together in small groups to collect the data and generate tables and graphs.

Set them the job of preparing the report as homework, over two or three nights or a weekend.

Have each group present their report to the class. Some reports should be presented orally (e.g. reports for illiterate livestock owners, briefings for busy decision-makers such as the Minister). Other reports should be written, and the contents and presentation can be explained.

### Questions for discussion

Did each report contain the information that was needed?

Was it easy to understand?

How could they be improved?

## Activity 22: Knowledge quiz competition

<b>Location:</b> Classroom	<b>Duration:</b> 30 minutes
<b>Reference:</b> -	<b>Software:</b> None
<b>Objectives</b> Encourage participation Fun break Warmer	<b>Concepts practiced</b> Any concept from the previous topics covered during the training
<b>Equipment and materials</b> Scoreboard (Blackboard/whiteboard/paper) Paper for writing questions	

### Description of activity

The group is divided into two. Each group is given 10 minutes to think of a list of 15 questions (with the right answers), based on the topics covered during the previous day, or throughout the training. After 10 minutes, a spokesperson for each group is chosen. The first group asks the first question. Group 2 has 30 seconds to answer (the trainer keeps time), during which they may discuss the answer amongst themselves. Group 1 says if the answer is correct or not. If the answer is correct, Group 2 gets a point. If the answer is wrong or they are unable to answer in the time, group 1 gets a point. If the answer is either right or wrong, but Group 1 judges it incorrectly, they lose a point.

It is then the second group's turn to ask a question. This continues until all questions have been asked. The group with the highest score wins.

### Questions for discussion

Ask individuals to provide the correct answer for any questions that were answered incorrectly. If nobody (including the asking group) is able to, set aside time to cover the area again.

# **Part IV**

## **Appendices**

**Appendix A: Glossary**

**Appendix B: Statistical Formulae**

**Appendix C: Computer Programs**

**Appendix D: Example Data Collection Forms**

**Appendix E: Contents of CD and Floppy Disks**



# Appendix A

## Glossary of Epidemiological Terms

(Courtesy of Dr Ian Gardner, University of California, Davis)

**Accuracy** - the degree to which a measurement, or an estimate based on measurements, represents the true value of the attribute that is being measured.

**Agent** - a factor such as a microorganism or chemical substance whose presence or excessive presence is necessary for the occurrence of a disease.

**Analytical study** - a hypothesis testing method of investigating the association between a given disease, health state, or other outcome variable, and possible causative factors.

**Benefit-Cost Ratio** - the ratio of the net present values (usually monetary values) of measurable benefits to costs. Used to determine the economic feasibility or probability of success of a time-bounded program.

**Bias** - any effect at any stage of an investigation tending to produce results that depart systematically from the true values i.e. a systematic error.

**Bias (Response bias)** - a systematic error due to differences in characteristics between those who volunteer to participate in a study and those who do not.

**Bias (Selection bias)** - error due to systematic differences in characteristics between those animals or herds which are selected for study and those which are not.

**Categorical Data** - qualitative data which can be allocated to specific groups. May be nominal (ie. named) or ordinal (ie. ordered) or dichotomous (ie. presence/absence).

**Chi-Square Test** - a method of testing to determine whether two or more series of proportions or frequencies are significantly different from one another or whether a single series of proportions differs significantly from an expected distribution. Pearson's Chi-square is used for unmatched data and McNemar's Chi-square for matched data. See definition of association for further explanation.

**Clustering** - a closely grouped series of events or cases of a disease in relation to time or place or both. The term is normally used to describe aggregation of relatively uncommon events or diseases.

**Confidence Limits** - an interval whose end points can be calculated from observational data and has a specified probability of containing the parameter of interest.

**Confounding** - a situation in which the effects of two factors are not separated. The distortion of the apparent effect of an exposure or risk factor brought about by association with other factors that can influence the outcome.

**Confounding Factor** - a confounding factor or variable is one which is distributed non-randomly with respect to the independent (exposure) variable and is associated with the dependent (outcome) variable being studied. The association with the dependent variable is usually established from results of previous studies.

**Contingency Table** - a tabular cross-classification of data such that subcategories of one characteristic are indicated horizontally (in rows) and subcategories of another characteristic are indicated vertically (in columns), and the number of units in each cell is indicated. The simplest contingency table is the fourfold or 2 x 2 table, but a contingency table may include several dimensions of classification.

**Continuous Data** - quantitative data with a potentially infinite number of possible values along a continuum.

**Cost Benefit Analysis** - methods of identifying the losses and gains in monetary terms of the effects of a disease that are incurred by society as a whole.

**Cross-Sectional Study** - (syn: prevalence study) - a study carried out on a representative sample of a population that examines the relationship between a disease or other health-related characteristic and other variables of interest as they exist in a defined population at one particular time.

**Crude Rate** - a rate which applies to a total population irrespective of the attributes of that population (cf. specific rate).

**Data** - facts of any kind. Data are plural, datum is singular.

**Data Base** - a systemized collection of information, commonly on electronic media about a specific subject such as animal disease.

**Denominator** - the population at risk in the calculation of a rate or ratio. See also Numerator

**Dependent Variable** - (syn:outcome/response variable) a variable or factor, the value of which depends on or is hypothesized to depend on the effect of other [causal] variable(s) in the study.

**Endemic Disease** - the constant presence of a disease or infectious agent within a given geographic area or population group. It also implies a prevalence which is usual in the area or in the population.

**Epidemic** - the occurrence in a population or region of cases of disease clearly in excess of normal expectancy - this is frequently taken as more than two standard deviations greater than the mean occurrence.

**Epidemic curve** - a histogram in which the X-axis represents the time of occurrence of disease cases and the Y-axis represents the frequency of disease cases. It is a useful tool to determine the epidemiology of disease occurrence in an outbreak investigation.

**Epidemic, Propagating** - an outbreak or series of outbreaks resulting from animal to animal spread.

**Epidemiology** - the study of the distribution and determinants of health related states and events in populations. It is a term now in common usage for studies in animal populations although epizootiology is still occasionally used.

**Epidemiology, Descriptive** - study of the occurrence of disease or other health related characteristics in populations. Implies general observation rather than analysis.

**Error, Sampling** - after testing a sample from a large population, the mean or any other statistic calculated from the sample will have a different value from the true value if the whole population was measured. The difference between the value for the whole population and its estimate calculated from the sample is called the sampling error.

**Error, Systematic** - that due to factors other than chance, such as faulty measuring instruments.

**False Negative** - when the result of an individual test is negative but the disease or condition is present.

**False Positive** - when the result of an individual test is positive but the disease or condition is not present.

**Frequency** - a count, or number of occurrences, of an event in a specified population and time period.

**Frequency Distribution** - any arrangement of numerical data obtained by measuring a parameter in a population.

**Histogram** - frequency distribution plotted in the form of rectangles whose bases are equal to the class width and whose areas are proportional to the absolute or relative frequencies.

**Hypotheses** - a proposition that can be tested by facts that are known or can be obtained. The assertion that an association between two, or more variables or a difference between 2 or more groups, exists in the larger population of interest.

**Incidence** - the number of new cases of disease or other condition which occur in a specified population during a given period. Mathematically, 2 types of incidence rate can be distinguished. These are incidence density rates and cumulative incidence.

**Incubation Period** - the interval of time between invasion by an infectious agent or contact with a chemical and the appearance of symptoms of the disease or condition in question.

**Independent Variable** - the characteristic being observed or measured that is hypothesized to influence an event. An independent variable is not influenced by the event or manifestation but may cause it or contribute to its variation.

**Index Case** - the first diagnosed case of an outbreak in a herd or other defined group.

**Infectivity** - the ability of an agent to enter, survive and multiply in the host. Epidemiologically, it is measured as the % of the individuals exposed to an agent who become infected.

**Inference** - the process of passing from observations to generalizations.

**Latent Infection** - persistence of an infectious agent within the host without symptoms of disease.

**Linear Regression** - statistical method used to study the relationship between independent and dependent variables when the dependent variable consists of continuous data.

**Longitudinal Study** - a study conducted over a defined period of time which may be either retrospective or prospective. See also Case Control and Cohort Study.

**Mean-Arithmetic** - a measure of central tendency computed by adding all the individual values together and dividing by the number in the group.

**Median** - the median is the middle value of a set of observations arranged in order of magnitude.

**Mode** - the mode is the most frequently occurring value in a set of observations. A given set of observations can have more than one mode. (see also Bimodal Distribution).

**Monitoring** - the performance and analysis of routine measurements aimed at the early detection of changes in the prevalence or incidence of disease, health, or alteration in a production parameter.

**Multistage Sampling** - a term applied to the selection of a sample in two or more stages. eg, selecting a sample of herds and then a sample of livestock within those herds.

**Nominal Data** - a type of data in which there are limited categories but no order, such as breed and eye color.

**Normal** - within the usual range of variation in a given population or population group; or frequently occurring in a given population or group.

**Normal Distribution** - a continuous symmetrical frequency distribution where both tails extend to infinity, the arithmetic mean, mode and median are identical. Graphically it is a bell shaped curve and its steepness or shape is completely determined by the mean and variance.

**Null Hypothesis** - the hypothesis that two variables have no association at all, or two or more population distributions do not differ from each other.

**Numerator** - the upper portion of a fraction used to calculate a rate or ratio.

**Observational Study** - an epidemiological study where nature is allowed to take its course while changes or differences in one characteristic are studied in relation to changes or differences in other(s) without intervention of the investigator (e.g. descriptive, cross-sectional case-control, cohort).

**Occurrence** - a statement indicating the presence of disease without signifying the frequency. This definition describes the use of the word in international animal disease reports.

**Ordinal data** - a type of data in which there are limited categories with an inherent ranking from lowest to highest (such as severity of disease).

**Outbreak** - the occurrence of disease in a herd or any other identifiable group of animals. For practical purposes, the term is synonymous with epidemic.

**Outliers** - observations differing so widely from the rest of the data as to lead one to suspect that a gross error in recording may have been committed, or suggesting that these values came from a different population.

**Pandemic** - an epidemic occurring over a very wide area, involving many countries and usually affecting a large proportion of the population.

**Parameter** - a summary descriptive characteristic of a population (cf statistic - which is a sample-based measure).

**Pathogenicity** - the ability of an organism to produce disease. Epidemiologically, it is measured as the % of infected individuals who develop clinical disease.

**Power** - probability of finding a difference between two or more groups given that a difference exists. Power = 1-Beta = 1-Probability of a type II error.

**Precision** - the quality of being sharply defined or stated. Refers to the ability of a test or measuring device to give consistent results when applied repeatedly. Sometimes also called repeatability.

**Predictive Value** - in screening or diagnostic tests, the predictive value of a positive test is the proportion of test positive animals that have the disease. The predictive value of a negative test is the probability that an animal with a negative test does not have the disease. The predictive value of a test is determined by the sensitivity and specificity of the test, and by the prevalence of the condition at the time the test is used.

**Prevalence** - the proportion of cases of a disease or other condition present in a population without any distinction between old and new cases. When used without qualification the term usually refers to the number of cases as a proportion of the population at risk at a specified point in time (point prevalence).

$$\text{Prevalence} = \frac{\text{No. cases at specific point in time}}{\text{Population at risk at same point in time}}$$

**Prevalence study** - see cross-sectional study

**Primary Case** - the individual that introduces disease into a herd, flock, or other group under study. Not necessarily the first diagnosed case in that herd. See index case.

**Proportion** - a fraction where the numerator is a subset of the denominator.

**Prospective Study** - see Cohort Study.

**Qualitative data** - that which possess specific qualities such as breed, gender, or color. See nominal data.

**Random** - governed by chance.

**Randomization** - allocation of individuals to groups by chance. Within the limits of chance variation, randomization should make control and experimental groups similar at the start of an investigation and ensure that personal judgement and prejudices of the investigator do not influence allocation. Note that random allocation follows a predetermined plan often devised with the aid of a table of random numbers or by an electronic random number generator.

**Random Sample** - a sample of a population assembled so that each member of the population has an equal and non-zero opportunity to be selected.

**Random Sampling** - procedure for selecting individuals from a population so that each has an equal chance of being selected in the sample.

**Rate** - an expression of the change in one quantity per unit time.

It is a ratio whose essential characteristic is that time is an element of the denominator and in which there is a distinct relationship between numerator and denominator. See also ratio and proportion.

**Ratio** - the expression of the relationship between a numerator and denominator where the two are separate and distinct quantities, i.e the numerator is not included in the denominator.

**Relative Risk** - the ratio of the disease incidence in individuals exposed to a hypothesized factor to the incidence in individuals not exposed; a measure of association commonly used in cohort studies. See also odds ratio.

	Diseased	Not diseased
Exposed	a	b
Unexposed	c	d

The Relative Risk is  $[a/(a+b)] / [c/(c+d)]$

**Repeatability** - the ability of a test to give consistent results in repeated tests. See precision.

**Response Rate** - the number of completed or returned survey instruments (questionnaires, interview etc.) divided by the total number of individuals selected for study.

**Retrospective Study** - a study that collects and utilizes historical data. A case-control study is retrospective because it looks back from the point of known effects to determine causative factors.

**Robust** - a statistical test is described as robust if the inferences hold true even when assumptions inherent in the tests are violated.

**Sampling** - the process of selecting a number of representative subjects from all the subjects in a particular group. Conclusions based on sample results may be attributed only to the population sampled. See also random sample and selection bias.

**Screening** - implies subjecting a population or sample of a population to a diagnostic test or procedure, with the objective of detecting disease. Tests used for this purpose are usually cheap, easily performed, sensitive but often not very specific.

**Sensitivity** - is the proportion of truly diseased animals in the screened population which are identified as diseased by the test. It is a measure of the probability that a diseased individual will be correctly identified by the test.

**Sentinel Herds** - herds that are reasonably representative of the population as a whole and which are tested at regular intervals for infectious disease to determine whether and to what extent the diseases are occurring in the population.

**Seroepidemiology** - epidemiological studies based on an examination of sera taken from the population or a sample of the population.

**Significance, Level of** - also known as alpha ( $\alpha$ ) or type I error rate. The probability of saying a difference exists when none does.

**Spatial distribution** - the relationship of disease events to location of individual animals or clusters of animals.

**Specificity** - is the proportion of truly non-diseased animals correctly identified by the test. Like sensitivity, specificity is a conditional probability.

**Specific Rate** - expresses the frequency of a characteristic per unit size of a specific population.

**Sporadic** - a disease occurring irregularly and generally infrequently and without any apparent underlying pattern.

**Standard Deviation** - a measure of dispersion or variation. Equal to the positive square root of the variance. The mean indicates where the values for a group are centered. The standard deviation is a measure of how widely values are dispersed around the mean in the population.

**Standard Error** - measure of the variability of a sample statistic that specifically relates an observed mean to the true mean of the population.

**Statistic** - a summary value calculated from a sample of observations usually to estimate a population parameter.

**Statistical Significance** - statistical methods allow an estimate to be made of the probability of the observed degree of association between independent and dependent variables being exceeded under a null hypothesis. From this estimate the statistical "significance" of a result can be stated. Usually the level of statistical significance is stated by the "P" value or probability value. See also Significance, Level of.

**Statistics** - the science and art of dealing with variation in data through collection, classification, and appropriate analysis.

**Stratified Sample** - involves dividing the population into distinct subgroups according to some important characteristic, eg herd size, and selecting a random sample out of each subgroup.

**Surveillance** - a system or measurement technique to gain knowledge about a population by collection, analysis, and interpretation of data with a view to the early

detection of cases of disease or changes in the health status of the population. The goal of surveillance is directed action in the treatment or prevention of the condition.

**Survey** - an investigation in which information is systematically collected.

**Systematic Sample** - the procedure of selecting according to some simple systematic rule, such as every 5th cow in the herd as they enter the milking parlor. A systematic sample may lead to errors that invalidate generalizations.

**Temporal Distribution** - the relationship of disease events to time.

**Trend** - a long-time movement in an ordered series (e.g. a time series). An essential feature is that the movement, whilst possibly irregular in the short term, shows movement consistently in the same direction over a long term.

**Type I Error** - an error which occurs when using data from a sample that demonstrates a statistically significant association when no such association is present in the population. Equals the level of significance or alpha.

**Type II Error** - an error that occurs from failure to demonstrate a statistically significant association when one exists in a population. Equals Beta. The power of a study equals 1-Beta.

**Validity** - the extent to which a study or test measures what it sets out to measure.

**Variable** - see Dependent variable, Independent variable.

**Variance** - the variance of a set of observations is the sum of squares of the deviation of each observation from the arithmetic mean of the observations, divided by one less than the number of observations.

**Virulence** - it is the degree of pathogenicity and indicates the potential severity of the disease produced by an agent in a given host. Epidemiologically, it is measured as the % of individuals with disease who become seriously ill or die. Sometimes, the case-fatality rate is considered an indicator for the virulence of disease.



# Appendix B

## Statistical Formulae

### Definition of symbols

The symbols used are based on the terminology of Levy and Lemeshow (1991) and Yamane (1967).

$M$  = number of clusters or PSUs in the population (villages).

$m$  = number of clusters sampled

$N$  = number of SSUs in the population (animals).

$n$  = number of SSUs sampled

$N_i$  = number of SSUs in the  $i$ th cluster

$n_i$  = number of SSUs sampled in the  $i$ th cluster

$\bar{n}$  = average number of SSUs sampled per cluster, which is equal to  $n_i$  if a constant number of SSUs is used

$\Pi$  = proportion of population with the characteristic of interest

$\pi_i$  = proportion of the  $i$ th cluster with the characteristic of interest

$s_{wtb}^2$  = sample variance between clusters

$s_{wti}^2$  = sample variance within clusters

$x_{ij}$  = the observation of the  $j$ th SSU in the  $i$ th cluster, coded as 1 for individuals with the characteristic of interest, and 0 for those without

$x_i$  = the sum of the  $j$  values in cluster  $i$ .

$V(\hat{\Pi})$  = variance of the estimated proportion

# Prevalence Surveys

## PPS Sampling

### Prevalence Estimate

The estimate of the proportion of the population with the characteristic of interest (for example, the proportion of animals with protective titres to FMD) (Yamane, 1967) is:

$$\hat{\Pi} = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} x_{ij}}{m \bar{n}} \quad (1)$$

The sampling scheme produces a self-weighting sample, which means that each SSU has an equal probability of being selected. An unbiased estimate of the population proportion is therefore simply the proportion of positive SSUs in the sample, or the total number of positive SSUs in the sample divided by the total sample size.

### Variance Estimate

An unbiased estimator of the variance of the estimate is given by:

$$V(\hat{\Pi}) = \frac{1}{m} \frac{1}{m-1} \sum_{i=1}^m (\hat{\pi}_i - \hat{\Pi})^2 \quad (2)$$

where  $\hat{\pi}_i$  is defined as:

$$\hat{\pi}_i = \frac{1}{\bar{n}} \sum_{j=1}^{\bar{n}} x_{ij} = \frac{x_i}{\bar{n}} \quad (3)$$

### Sample Size Calculation

The usual method of calculating the optimal value of  $m$  and  $\bar{n}$  is based on the use of a cost function (e.g. Levy and Lemeshow, 1991 p262). The choice of cost function depends on the nature of the survey work, and can be quite complex. For instance, the curve produced may not be continuous. This would occur if, at some point, the number of animals per village exceeded that which could be examined in one day. After this point, extra travel or accommodation costs may be incurred. The nature of travel costs also depend on the way in which field sites (clusters) are visited: they may be visited one at a time, returning to a central base each time, or two or more may be visited on each trip. Despite these complications, a simple cost function will usually be adequate, taking the general form:

$$C = C_0 + C_1 m + C_2 m \bar{n} \quad (4)$$

where  $C$  = total costs,  $C_0$  = fixed costs,  $C_1$  = cost per PSU, and  $C_2$  = cost per SSU. The value of  $\bar{n}$  which will minimise the variance subject to the constraint of the cost function can be found by (Yamane, 1967):

$$\bar{n} = \sqrt{\frac{C_1 \frac{S_{wti}^2}{N}}{C_2 \left( S_{wtb}^2 - \left( \frac{S_{wti}^2}{N} \right) \right)}} \tag{5}$$

If the number of animals sampled per village is much less than the village population, this simplifies to:

$$\bar{n} = \sqrt{\frac{C_1 \frac{S_{wti}^2}{N}}{C_2 S_{wtb}^2}} \tag{6}$$

where the sample estimate of the within-cluster variances is:

$$s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \frac{\bar{n} \hat{\pi}_i (1 - \hat{\pi}_i)}{\bar{n} - 1} \tag{7}$$

the mean of the within-cluster variances is:

$$\bar{s}_i^2 = \frac{1}{m} \sum_{i=1}^m s_i^2 = s_{wti}^2 \tag{8}$$

which is equal to the weighted mean of the within-cluster variances, as the weights  $n_i$  are all equal. The sample estimate of the variation between clusters is:

$$s_{wtb}^2 = \frac{1}{m - 1} \sum_{i=1}^m (\hat{\pi}_i - \hat{\Pi})^2 \tag{9}$$

Formula 5 can be used to calculate the optimal number of animals (resulting in minimum total variance) to be sampled per village for specified per village and per animal costs, and variance estimates. When a fixed budget is available for the survey, the optimal number of villages required for a survey of a given cost can be calculated by substituting this value of  $\bar{n}$  into the cost function 4. For ongoing surveillance, ensuring that survey estimates achieve the necessary level of precision is more important. The number of villages that yield an estimate of a given accuracy may be calculated in the following way.

The confidence interval for the prevalence estimate is given by (Levy and Lemeshow, 1991 p53):

$$CI = \hat{\Pi} \pm z_{(1-\frac{\alpha}{2})} \sqrt{V(\hat{\Pi})} \tag{10}$$

where  $z_{(1-\frac{\alpha}{2})}$  is the standard normal deviate (1.96 for a 95% confidence interval). The variance Formula 2 can be rewritten as:

$$V(\hat{\Pi}) = \frac{s_{wtb}^2}{m} \quad (11)$$

By substituting Formula 9 into Formula 2 and rearranging, we get

$$m = \frac{s_{wtb}^2}{\left(\frac{u}{z_{(1-\frac{\alpha}{2})}}\right)^2} \quad (12)$$

where  $u$  is half the width of the confidence interval. This approach is adequate when the estimated proportion is neither very large or very small. However, if the prevalence is high or low, then using a fixed-width target confidence interval may be inappropriate. An alternative approach involves the use of relative error,  $R$ , defined as:

$$R = \frac{\sqrt{\text{Var}(\hat{\Pi})}}{\hat{\Pi}} = \frac{s_{wtb}}{\sqrt{m} \hat{\Pi}} \quad (13)$$

In order to ensure that the desired relative error is achieved for proportions down to  $\Pi_0$ , the appropriate number of villages to sample using this approach can then be calculated as:

$$m = \frac{s_{wtb}^2}{\Pi_0^2 R^2} \quad (14)$$

### SRS Sampling

PPS sampling provides estimates of relatively low variance and the selection of a fixed number of animals per village makes field work more predictable. However to achieve these benefits, a sampling frame which includes reliable data on village livestock populations is required. When a sampling frame with no such data available, simple random sampling (SRS) must be used at the first stage. In order to achieve a self-weighting sample (in which every animal in the population has the same probability of selection), a fixed proportion of the village population must be sampled at the second stage. This sampling scheme therefore requires simple random sampling with replacement of  $m$  villages from a total of  $M$ , and then simple random sampling (without replacement) of  $n_i$  animals from the  $i$ th village total of  $N_i$ , such that  $n_i/N_i$  (the second-stage sampling fraction, or  $f_2$ ) is constant (or nearly so).

### Prevalence Estimate

As  $N$ , the total number of animals in the population, is unknown, this value must be estimated. Using the ratio-to-size estimate (Cochran, 1977 p303), the estimated proportion is:

$$\hat{\Pi} = \bar{x} = \frac{\frac{M}{m} \sum_{i=1}^m \frac{N_i}{n_i} \sum_{j=1}^{n_i} x_{ij}}{\frac{M}{m} \sum_{i=1}^m N_i} = \frac{\sum_{i=1}^m N_i \bar{x}_i}{\sum_{i=1}^m N_i} \tag{15}$$

This mean per village is equal to the mean per animal when  $f_2$  is constant. Ratio estimates of this nature are subject to some bias, but are required when an estimate of the population size is not available.

**Variance Estimate**

When biased estimators are used, estimates of the mean square error (mse) are preferable to the variance, as they take this bias into account. An estimator of the mse for the above ratio estimate is given by:

$$var(\bar{x}) = mse(\hat{\Pi}) = \frac{1}{\hat{N}^2} \left[ \frac{M^2}{m} \left(1 - \frac{m}{M}\right) \frac{\sum_{i=1}^m N_i^2 (\bar{x}_i - \bar{x})^2}{m-1} + \frac{M}{m} \sum_{i=1}^m \frac{N_i^2 \left(1 - \frac{n_i}{N_i}\right) s_{2i}^2}{n_i} \right] \tag{16}$$

**Sample Size Calculation**

For optimal allocation of number of villages and number of animals based on costs assuming the same cost function as formula 4 above, except that the per-village costs will now include the cost of developing a sampling frame for the village. The optimal average number of animals sampled per village ( $\bar{n}$ ) can be calculated (Cochran, 1977 p314):

$$\bar{n}_{opt} = \frac{S_2}{\sqrt{S_b^2 - \frac{S_2^2}{\bar{N}}}} \sqrt{\frac{c_1}{c_2}} \tag{17}$$

where the weighted variance among villages per animal, and a slightly biased (upwards) estimate are:

$$S_b^2 = \frac{\sum_{i=1}^M N_i^2 (\bar{X}_i - \bar{\bar{X}})^2}{\bar{N}^2(M-1)} \quad \hat{S}_b^2 = \frac{\frac{M}{m} \sum_{i=1}^m N_i^2 (\bar{x}_i - \bar{\bar{x}})^2}{\bar{N}^2(M-1)} \tag{18}$$

and the weighted mean of within village variances and its unbiased sample estimate are:

$$S_2^2 = \sum_{i=1}^M \frac{N_i}{N} S_{2i}^2 \quad \hat{S}_2^2 = \frac{M}{m} \sum_{i=1}^m \frac{N_i}{N} s_{2i}^2 \tag{19}$$

Calculation of the optimal second-stage sampling fraction ( $f_2$ ) requires an estimate of the average village livestock population:

$$f_2 = \frac{\bar{n}}{N} \quad (20)$$

Calculation of the optimal number of villages to sample is based on the following formula:

$$mse(\hat{\Pi}) = \frac{1}{m} \left( S_b^2 - \frac{S_2^2}{N} + \frac{1}{n} S_2^2 \right) - \frac{S_b^2}{N} \quad (21)$$

To calculate the optimal number of villages based on a fixed width confidence interval (size  $2 \times u$ ),

$$m_{opt} = \frac{S_b^2 - \frac{S_2^2}{N} + \frac{1}{n} S_2^2}{\left( \frac{u}{z_{(1-\frac{\alpha}{2})}} \right)^2 + \frac{S_b^2}{N}} \quad (22)$$

Alternatively, the relative error can be used as in formula 14 above:

$$m_{opt} = \frac{S_b^2 - \frac{S_2^2}{N} + \frac{1}{n} S_2^2}{\Pi_0^2 R^2 + \frac{S_b^2}{N}} \quad (23)$$

## RGCS Sampling

Random geographical coordinate sampling (RGCS), described in Chapter 3, allows the random selection of villages in the absence of a village sampling frame. When using RGCS for the first stage of a two-stage sampling scheme, calculation of an unbiased estimate requires village proportions to be weighted by the number of villages within the selection radius for the randomly selected point.

### Prevalence Estimate

The estimate of the proportion is:

$$\hat{\Pi} = \frac{\sum_{i=1}^m N_i \bar{x}_i w_i}{\sum_{i=1}^m N_i w_i} \quad (24)$$

where  $m$  is the total number of villages and  $w_i$  is the number of villages around the  $i$ th random point. If the proportion of circles which straddle the border of the study region is large (say greater than 10%) it is advisable to modify Equation 24 to take the area of the circles outside the study region into account. This is achieved by replacing  $w_i$  with  $w_i/c_i$ , where  $c_i$  is the proportion of the  $i$ th circle lying within the study region.

**Variance Estimate**

An estimator for the variance of the estimated proportion is

$$var(\hat{\Pi}) = \left( \frac{1}{m(m-1)\hat{N}^2} \right) \left( \frac{A}{\pi r^2} \right)^2 \sum_{i=1}^m N_i \frac{w_i}{c_i} (\pi_i - \hat{\pi})^2 \tag{25}$$

where  $\hat{N}$  is the estimated total number of animals in the population, A is the total area of the study region and r is the selection radius used.  $\hat{N}$  may be estimated as:

$$\hat{N} = \frac{A}{a} \sum_{i=1}^m N_i w_i \tag{26}$$

where a is the total area inside the selection radii around the random coordinates. This is equal to  $(m_t \pi r^2) - a_{\text{external}}$ , where  $m_t$  is the total number of circles used (including those with no villages), r is the selection radius, and  $a_{\text{external}}$  is the sum of the area of the parts of circles lying outside the study area. These values are most easily calculated with a GIS.

The variance given by Formula 25 may be biased if the variability of the quantity of interest is greater in high density areas than in lower density areas. If there is reason to suspect that this is the case, then simple random sampling would be preferable. However, when no sampling frame can be constructed, RGCS may be the only alternative despite this potential bias in the variance.

**Sample Size Calculation**

The optimal second stage sampling fraction is found from:

$$f_2 = \sqrt{\frac{B_2 C_1}{(B_1 - B_2) C_2 \bar{N}'}} \tag{27}$$

where the values and their corresponding estimates are:

$$\begin{aligned} B_1 &= \sum_{i=1}^M N_i^2 \frac{w_i}{c_i} (\hat{\pi}_i - \hat{\pi})^2 & \hat{B}_1 &= \frac{A}{a} \sum_{i=1}^m N_i^2 \frac{w_i}{c_i} (\hat{\pi}_i - \hat{\pi})^2 \\ B_2 &= \sum_{i=1}^M N_i \frac{w_i}{c_i} S_{2i}^2 & \hat{B}_2 &= \frac{A}{a} \sum_{i=1}^m N_i \frac{w_i}{c_i} \hat{S}_{2i}^2 \\ C_1 &= C_u + C_l \bar{N}' & \hat{S}_{2i}^2 &= \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{ij} - \hat{\pi}_i)^2 \\ \bar{N}' &= \frac{1}{W} \sum_{i=1}^M N_i \frac{w_i}{c_i} & \hat{\bar{N}}' &= \frac{1}{W} \frac{A}{a} \sum_{i=1}^m N_i \frac{c_i}{w_i} \\ W &= \frac{A}{\pi r^2} \end{aligned} \tag{28}$$

The approximate relationship between  $B_1, B_2, \hat{\bar{N}}'$  and  $s_b^2, s_2^2$  and  $\bar{N}$  (using average weighting values) is as follows:

$$\begin{aligned}\hat{B}_1 &\approx \bar{N}^2 (m-1) \frac{A}{a} \hat{S}_b^2 \left( \frac{w_i}{c_i} \right) \\ \hat{B}_2 &\approx \frac{1}{N} \hat{S}_2^2 \left( \frac{w_i}{c_i} \right) \\ \hat{N}' &\approx \left( \frac{c_i}{w_i} \right) \bar{N}\end{aligned}\quad (29)$$

To calculate  $m$ , these equations can be combined with the desired confidence interval half-width,  $u$ , or desired relative error,  $R$ , using:

$$V = \frac{u^2}{z_{1-\frac{\alpha}{2}}^2} \quad \text{or} \quad V = R^2 \Pi_0^2 \quad (30)$$

The number of villages to be sampled is then found from:

$$m = \left( \frac{W}{N^2 V} \right) \left[ B_1 - B_2 + \frac{B_2}{f_2} \right] \quad (31)$$

### Stratified Sampling

Stratification of the sample will almost always improve the variance of the estimate (Levy and Lemeshow, 1991 p105), as well as providing some practical advantages. Stratification is also important if separate estimates need to be made for the different strata. The OIE recommends stratification for random geographical coordinate sampling for these reasons (OIE, 1990). Ideally strata should be created such that the variation within strata is minimised, and variation between strata is maximised. However, for surveys of this type, very little population data is available. Stratification by geographical area is therefore perhaps the only practical option. It can generally be expected that serological response will be variable between regions, and this stratification would lead to lower variance. Administrative subdivisions are the most readily available geographical areas for stratification.

A conceptually simple approach to stratified sampling is to use proportionate sampling of elements. This means that within each stratum, the sampling fraction is equal to the overall sampling fraction for the whole sample. Each stratum's contribution will be proportional to its size.

Under stratified sampling, each of the strata can be treated as independent samples, and the proportions and standard errors calculated separately. To combine these strata estimates into an overall estimate, each stratum is assigned a weight,  $W_h$ . These weights are generally the proportion of the population contained within the stratum. In this case, an estimate of the proportion of villages in each stratum is used, as the number of livestock in each stratum is usually unknown. Where it is impossible to estimate the number of villages in a stratum, the area of the geographical subdivision could be used to weight strata. Kish (1995 p80) points out that when estimates are used as weights, the formulae used to combine strata no longer strictly hold, and will simply be approximations. In practice, good estimates of the number of villages in each area will often be available.

Under proportionate sampling, a constant sampling fraction is used, so that (Kish, 1995 p80)

$$f = \frac{n_h}{N_h} = \frac{n}{N} \quad (32)$$

where  $f$  is the sampling fraction,  $n_h$  is the number of villages chosen per district,  $N_h$  is the total number of villages per district,  $n$  is the sample size and  $N$  is the population total number of villages. In this case, the stratum weights will be

$$W_h = \frac{n_h}{n} = \frac{N_h}{N} \quad (33)$$

The estimate of the population proportion,  $P_t$ , across strata is then simply the weighted sum of the strata proportions, using the calculated weights:

$$P_t = \sum W_h P_h \quad (34)$$

The estimate of the variance is equal to the sum of the strata variances weighted by the square of the strata weights:

$$\text{var}(P_t) = \sum W_h^2 \text{var}(P_h) \quad (35)$$

Calculation of the final estimates is therefore simply a matter of calculating the strata estimates and combining them with the above formulae. The strata estimates may be derived in any way, but in this case are all calculated based on the two-stage sampling design used to collect them.

When a ratio estimate is being used (as in the formulae above for simple random sampling and random geographic coordinate sampling), and either the number of villages in each stratum is small, or the number of strata is large, this approach may lead to significant bias in the estimates. In this case, the combined ratio estimate  $\hat{\Pi}_c$  is preferable (Cochran, 1977 p320):

$$\hat{\Pi}_c = \frac{\sum_{h=1}^L \hat{X}_h}{\sum_{h=1}^L \hat{N}_h} \quad (36)$$

An estimator of the variance of the combined estimate is:

$$v(\hat{\Pi}_c) = \frac{1}{N^2} \sum_{h=1}^L \frac{1}{m_h(m_h-1)} \sum_{i=1}^{m_h} (d'_{hi} - \bar{d}'_h)^2 \quad (37)$$

where

$$\begin{aligned}
 d'_{hi} &= \frac{N_{hi} \bar{d}_{hi}}{Z_{hi}} \\
 \bar{d}'_h &= \frac{1}{m_h} \sum_{i=1}^{m_h} d_{hi} \\
 \bar{d}_{hi} &= \bar{x}_{hi} - \hat{\Pi}_c
 \end{aligned} \tag{38}$$

and the selection probabilities are:

$$\begin{aligned}
 Z_{hi} &= \frac{1}{M_h} && \text{for simple random sampling} \\
 &= \frac{\pi r^2}{A} \frac{c_i}{w_i} && \text{for random geographic coordinate sampling.}
 \end{aligned} \tag{39}$$

## Apparent Prevalence to True Prevalence

### Prevalence Calculation

Prevalence estimates based on the use of an imperfect test must be corrected to take account of test performance. The formula to convert Apparent Prevalence (AP) to True Prevalence is:

$$\text{True Prevalence} = \frac{\text{AP} + \text{Sp} - 1}{\text{Se} + \text{Sp} - 1} \tag{40}$$

### Confidence Interval

If the prevalence estimate has been calculated using simple random sampling, the confidence interval can be calculated from the variance estimate, given by:

$$\text{var}(\hat{p}) = \frac{p(1-p)}{n(\text{Se} + \text{Sp})^2} \tag{41}$$

The confidence interval is

$$(\hat{p} - (Z_{\frac{\alpha}{2}} \times \sqrt{\text{var}(\hat{p})}), \hat{p} + (Z_{\frac{\alpha}{2}} \times \sqrt{\text{var}(\hat{p})})) \tag{42}$$

# Incidence Rate Surveys

## Capture / Recapture

### Estimate of Total

Seber (1970) uses an unbiased modification of the original formula shown in the text:

$$\hat{N} = \frac{(n_A + 1)(n_B + 1)}{(n_{11} + 1)} - 1 \quad (43)$$

### Confidence Interval

The variance can be calculated as (Seber, 1970)

$$\text{var}(\hat{N}) = \frac{(n_A + 1)(n_B + 1)(n_A - n_{11})(n_B - n_{11})}{(n_{11} + 1)^2(n_{11} + 2)} \quad (44)$$

McCarty *et al* (1993) calculate the 95% confidence interval by assuming a normal distribution and using  $\hat{N} \pm 1.96\sqrt{\text{var}}$ . Yip *et al* (1995) warn that this approach is not reliable for log-linear modelling, and that a likelihood interval (Hook and Regal, 1982) or bootstrap procedure (Buckland, 1984) should be used. In this two-sample example, the variance Formula 44 has been used and a normal distribution assumed.

## Survival Analysis

### Sample Size

The sample size required to detect a difference between two populations may be calculated using the following formula (Lee 1992 p 341). The formula indicates the number of uncensored observations required in each group.

$$n_d = \frac{2\tau(1, \alpha, \beta)}{(\log_e a)^2} \quad (45)$$

Where  $a = \frac{\mu_1}{\mu_2}$  and  $\mu_1$  is the larger of the expected median survival times of the two groups and  $\mu_2$  is the smaller and  $\tau(1, \alpha, \beta)$  is a non-centrality parameter for 2 groups, with alpha and beta being type I and type II errors.

### Log-Rank Test

The formula for the chi-square test used to determine if the survival experience of two populations is different is (Lee 1992)

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \quad (46)$$

### Hazard Ratio

The hazard ratio,  $R$ , is a measure of the relative survival experience of the two groups (Altman, 1991). It is defined as:

$$R = \frac{\left(\frac{O_1}{E_1}\right)}{\left(\frac{O_2}{E_2}\right)} \quad (47)$$

where  $O$  represents the observed number of failures and  $E$  the expected number of failures, and the subscripts represent groups 1 and 2. Altman (1991) also presents an alternative approximation for the log hazard ratio, based on the same variance formula as used in the log-rank test which allows the calculation of approximate confidence intervals for the ratio.

# Freedom From Disease

## Probability Formula

In this discussion, the meaning of the symbols used is as follows:

- p prevalence
- Se sensitivity;
- Sp specificity
- D<sup>+</sup> disease<sup>1</sup> positive animals (true positives);
- D<sup>-</sup> disease negative animals (true negatives)
- T<sup>+</sup> test positive animals (positive reactors);
- T<sup>-</sup> test negative animals (negative reactors)
- P() the probability of an event with the event of interest described in the brackets
- x the number of T<sup>+</sup> in a sample
- y the number of D<sup>+</sup> in a sample
- n sample size
- N population size
- d number of diseased D<sup>+</sup> animals in the population
- $\binom{n}{x}$  the number of ways that x objects can be drawn from n, equal to  $\frac{n!}{x!(n-x)!}$

### Infinite population (or sampling with replacement)

The probability of observing x reactors when testing n animals from an infinite population is given by the binomial distribution modified to take account of test sensitivity and specificity:

$$P(T^+ = x) = \binom{n}{x} [pSe + (1 - p)(1 - Sp)]^x [p(1 - Se) + (1 - p)Sp]^{n-x} \quad (48)$$

### Finite population (sampling without replacement)

To overcome the limitations of other commonly used formulae (the assumption of an infinite population or sampling with replacement), the hypergeometric distribution can be modified for imperfect tests. The number of D<sup>+</sup> in the sample has a hypergeometric distribution. Given y D<sup>+</sup> in the sample, the number of true positives is Bin(y, Se), and the number of false positives is Bin(n-y, 1-Sp). We will have x T<sup>+</sup> if we have j true positives and x-j false positives. By considering the possible values of y and j, we can write down:

$$P(T^+ = x) = \sum_{y=0}^d \frac{\binom{d}{y} \binom{N-d}{n-y}}{\binom{N}{n}} \sum_{j=0}^{\min(x,y)} \binom{y}{j} Se^j (1-Se)^{y-j} \binom{n-y}{x-j} (1-Sp)^{x-j} Sp^{n-x-y+j} \quad (49)$$

---

<sup>1</sup>Disease in this context is defined in its broadest context: possessing the abnormality or state of interest. In surrogate tests for disease, it may mean, for example, the presence of antibodies.

## References

- Altman, D.G. (1991) Analysis of survival times. In: *Practical Statistics for Medical Research*. Chapman & Hall, London, 365-395
- Buckland, S.T., (1984) Monte Carlo confidence intervals. *Biometrics* **40**:811-817
- Cochran, W.G. (1977) *Sampling Techniques*, 3rd ed. John Wiley & Sons, Inc., New York
- Hook, E.B. and Regal, R.R., (1982) Validity of Bernoulli census, log-linear, and truncated binomial models for correcting for underestimates in prevalence studies. *American Journal of Epidemiology* **116**:168-176
- Kish, L. (1995) *Survey Sampling*. John Wiley and Sons, New York
- Lee, E.T. (1992) *Statistical methods for survival data analysis*, 2nd ed. John Wiley & Sons, Inc, New York
- Levy, P.S. and Lemeshow, S. (1991) *Sampling of Populations: Methods and Applications*, 2nd ed. John Wiley & Sons, Inc, New York
- McCarty, D.J., Tull, E.S., Moy, C.S., Kwoh, C.K. and LaPorte, R.E., (1993) Ascertainment corrected rates: Applications of capture-recapture methods. *International Journal of Epidemiology* **22**:559-565
- OIE (1990) *Guide to epidemiological surveillance for rinderpest*. Office International des Épizooties, Paris
- Seber, G.A.F., (1970) The effect of trap response on tag recapture estimates. *Biometrics* **26**:13-22
- Yamane, T. (1967) Cluster Sampling (II):Probability proportional to size. In: *Elementary Sampling Theory*. Prentice-Hall, Englewood Cliffs, N.J., 237-271
- Yip, P.S.F., Bruno, G., Tajima, N., Seber, G.A.F., Buckland, S.T., Cormack, R.M., Unwin, N., Chang, Y.-F., Fienberg, S.E., Junker, B.W., LaPorte, R.E., Libman, I.M. and McCarty, D.J., (1995) Capture-recapture and multiple record systems estimation II: Applications in human diseases. *American Journal of Epidemiology* **142**:1059-1068

# **Appendix C**

## **Computer Programs**

## Random Village

---

**Purpose**

Selection of a random sample from a sampling frame, using simple random sampling (SRS) or probability proportional to size sampling (PPS), optional replacement and stratification.

**Input**

- Data file in dBASE or Paradox format
- File should contain identifier for each element, and optionally, fields for size (population) and stratification.

**Output**

- List of randomly selected elements
- May be printed or saved as a new table

**Dos Version?** No

**Page Reference:** 52

## Random Animal

---

**Purpose**

Simple random sampling of animals from a village livestock sampling frame.

**Input**

- Number of animals owned by each livestock owner

**Output**

- Randomly selected animals identified by ID number of livestock owner, and sequential animal number.

**Dos Version?** Yes

**Page Reference** 60

**RGCS (Win95)**

---

**Purpose**

Selection of random coordinates for random geographic coordinate sampling

**Input**

- Coordinates of a rectangle bounding the study area (Cartesian or Decimal degrees format)
- Number of points to select

**Output**

- Random coordinates
- May be printed or saved to a new table

**Dos Version?** No

**Page Reference** 70

**RGCS (ArcView GIS)**

---

**Purpose**

Selection of random coordinates within one or more irregular polygons for random geographic coordinate sampling.

**Input**

- Digital map (ArcInfo coverage or ArcView Shapefile format) with polygon showing study area.
- Number of points to select
- Polygons representing the study area
- Selection radius

**Output**

- Random coordinates, displayed on screen with bounding circles determined by the selection radius
- Coordinates may be printed or saved to database file.
- Map display may be manipulated or superimposed over remotely sensed data to screen selected points.

**Dos Version?** No

**Page Reference** 71

## Prevalence

---

### Purpose

Calculation of sample sizes for two stage prevalence surveys using one of three survey designs (SRS, PPS, RGCS), and calculation of prevalence and other estimates from the results of such surveys.

### Input

#### For analysis of survey data:

- Data file in Paradox or dBASE format with disease status, first stage sampling unit (village) ID.
- Optionally, (depending on survey design) a second file with stratum ID, village population, weights, selection radius, and the size of the study area.

#### For sample size calculation:

- Estimated prevalence
- Within- and between-village variance estimates
- Cost per village and cost per animal
- Total number of villages and average village population
- Confidence level and desired accuracy

### Output

#### Data analysis:

- Estimate of prevalence, with variance and confidence interval.

#### Sample size calculation:

- Optimal (minimum cost) first and second stage sample sizes.

**Dos Version?** No

**Page Reference** 162

## Compare Prevalence

---

### Purpose

Compare prevalence estimates from two surveys to determine if the difference is likely to be real or just due to chance.

### Input

- Prevalence and variance estimates from two surveys.

### Output

- P value: Probability that the two observations came from the same population.

**Dos Version?** No

**Page Reference** 170

## True Prevalence

---

**Purpose**

Calculate the true prevalence based on the apparent prevalence and test performance.

**Input**

- Apparent prevalence
- Test sensitivity
- Test specificity

**Output**

- True prevalence
- Confidence interval (assuming single stage simple random sampling)

**Dos Version?** Yes

**Page Reference** 168

## Survival

---

**Purpose**

Perform survival analysis. Specifically to analyse data from retrospective disease outbreak surveys, create a Kaplan-Meier survival curve and statistics.

**Input**

- Data file in dBASE or Paradox format, with time, censoring and optionally weight and grouping fields.

**Output****Single group analysis:**

- Kaplan-Meier survival curve
- Mean and median survival time

**Two-group analysis**

- Kaplan-Meier survival curve
- Log rank test statistic and P value
- Hazard ratio and confidence interval

**Dos Version?** Yes

**Page Reference** 181

## Survive Size

---

### Purpose

Sample size calculation for survival analysis

### Input

- Mean or median survival times for two groups representing the minimum difference that can be distinguished

### Output

- Number of non-censored observations (villages recalling outbreaks) required for each group.

**Dos Version?** No

**Page Reference** 176

## CapRecap

---

### Purpose

Calculate population total based on two data sources

### Input

- Total number in data source 1
- Total number in data source 2
- Total appearing in both data sources

### Output

- Estimate of total in population with confidence interval

**Dos Version?** Yes

**Page Reference** 187

## FreeCalc

---

### Purpose

Calculate sample sizes for surveys to demonstrate freedom from disease, and analyse the results of such surveys

### Input

- Test sensitivity
- Test specificity
- Population size
- Minimum expected (maximum acceptable) prevalence
- Type I and II error levels
- Additionally, for analysis of results, sample size and number of positive reactors

### Output

- Sample size and cutpoint number of reactors
- Probability that the population is diseased

**Dos Version?** Yes

**Page Reference** 204



# Appendix D

## Example Data Collection Forms

The forms on the following pages can be copied or used as a model for developing your own data collection forms for a survey. In addition to the information shown, all sheets should have a few lines at the top saying where the data came from, who it was collected by, and when.

### **Village Livestock Sampling Frame**

This sheet is used during village interview to record the number of animals kept by each of the livestock owners. If more than one species is being sampled, extra columns can be added. Extra copies of this sheet (without pre-printed numbers) should be available for larger villages. When using them, make sure to number the livestock owners on the extra sheets sequentially, and keep the sheets together.

### **Random Number Table**

This table can be used for any random selection exercise, but in particular for selecting animals from the sampling frame made during the village interview;

### **Specimen Collection Sheet**

The species column is only necessary for multiple species surveys. Extra columns may be added for other relevant information (eg vaccination history).

### **Disease Outbreak Questionnaire**

This form was designed for FMD. Substitute the disease of interest.

### **Disease Ranking and Seasonal Patterns**

Be sure to accurately record a description of the disease.

### **RGCS Village Sheet (Sheet 1) and RGCS Random Point Sheet (Sheet 2)**

Sheets to be used for random geographic coordinate sampling fieldwork.

## Village Livestock Sampling Frame

No	Name of Livestock Owner	Total Animals	Cumulative Total	Selected Animals
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				
31				
32				
33				
34				
35				
36				
37				
38				

## Random Number Table

9537	7654	2531	7467	2873	5885	5154	6419	9346	9458	2281	4520	1241	6730	4263
3014	7669	2948	7241	0139	3841	1369	1123	8300	7790	3632	9154	4698	3874	2423
2682	4082	3359	0932	6215	9668	4282	7428	2833	7014	0217	2737	6768	4218	3007
5531	1283	5400	7610	3466	2697	0649	2159	4803	7655	3325	7537	5885	1465	4746
4534	4703	1566	8974	8989	3953	5752	4976	1253	1041	2678	0067	1001	1802	8224
4202	9222	0395	0882	0406	5696	4204	7995	0571	0744	6751	8284	7202	2610	2531
4783	0798	7713	5203	3246	9008	1017	6802	5738	9416	0092	3831	4662	7819	5152
1515	3328	4102	2777	3867	8974	0632	1175	6051	8063	2795	5037	2319	6941	0285
7824	5298	1243	0754	4284	9480	4027	6284	1251	7275	9796	9015	0199	7321	3200
3894	3231	2288	0103	7834	2159	6589	7655	4435	2457	0141	3600	6792	1631	0840
4495	1477	3933	1570	7080	6521	1885	5664	2691	7577	8866	2425	0383	5134	1282
2495	0365	0326	0856	7851	0801	9001	7861	6828	4483	6681	8913	5735	9767	7244
6941	7266	1482	6315	5838	5539	3608	9895	4136	7294	5075	7471	0057	4551	1275
7136	7584	1352	4940	4637	4448	5390	8329	0559	3921	7029	2652	4622	4366	2786
6602	5200	3213	4913	6662	9579	7025	1113	1206	9229	5973	9585	0994	1648	9597
3346	4427	2525	5519	0821	0334	2335	4005	0598	6894	8161	1447	3213	7990	9132
5327	7977	9909	7696	3362	8331	3798	3732	6549	9457	6097	2249	9890	5228	6924
2541	7991	9425	0987	0809	2695	2051	1145	4111	8633	3193	5735	2601	8008	2604
9611	9655	9767	5203	6374	2752	2562	0175	8457	0393	2300	3658	9471	2385	6007
5322	9436	8575	7562	3770	7711	7100	0856	8138	1847	3270	9227	5393	7474	8566
7959	2467	2482	8581	4816	5323	0199	7210	2602	9477	7211	4004	2738	9695	7642
7906	6113	8081	2517	9752	4073	3221	3255	0388	0730	7586	9013	9009	1631	3952
1374	9257	1451	0624	1662	5929	1230	2935	6900	3504	0815	3387	5632	0377	4424
1676	9319	6404	8020	8916	9174	0284	2252	3169	0590	1531	6276	1788	3408	6972
6970	1559	4110	7432	2041	3362	5336	4365	9501	8548	0159	0352	4491	4694	4804
5850	7679	9254	5612	3905	0924	1378	0962	0437	3103	2957	7646	5019	2527	1399
4712	3274	0387	0697	4663	2449	3002	5661	9899	5543	7188	1043	6954	0520	5805
3291	6142	4611	1300	5324	5192	0015	7741	7972	7192	6577	7169	8827	3935	9888
7277	9996	9284	0611	6375	6807	9284	6975	3175	1465	4700	8996	3251	8478	7923
9425	0618	5866	1284	0362	8875	5458	2846	6681	5532	6480	8909	7075	4222	1831
3045	3952	3590	9404	9828	7222	5711	3926	7353	6153	0426	5545	9608	9806	7823
4299	8225	3096	8302	4524	8587	6188	5714	9020	6674	6780	0167	8418	4586	2754
2207	4564	2702	5504	4287	5653	0294	7690	3897	4751	9238	0857	4756	8867	0935
7750	3178	2451	8603	6500	4976	1476	2884	8548	2806	0380	5326	3127	4905	4731
6009	4643	3594	8319	9547	4857	5677	5734	1317	5770	3484	5591	2051	3796	4675
7711	8280	3680	9546	6147	2663	1095	6521	2602	3125	5871	0333	5523	5951	7422
9115	2208	9888	3651	2995	3651	5409	3153	1912	4784	1442	3188	7233	5272	2297
1634	2060	5774	7820	5607	5813	3150	3583	8092	2846	2552	7785	2049	9719	9730
5092	7923	9073	9726	9775	7783	8331	4648	1630	3745	3901	2776	1808	3408	7362
1041	1523	0736	8295	3543	9323	0040	5601	0440	7831	3570	2664	4956	7887	2088
5022	8169	7826	3863	6097	6440	1104	7124	3058	5921	8873	2708	2044	2776	8838
9198	0531	5469	3493	2502	5640	2531	9095	5617	4837	7192	8672	1628	8392	9365
9246	3728	5474	9748	5657	4377	8841	2910	1538	6470	4421	4721	3605	5547	6820
1925	3806	1808	3684	9405	7201	1973	6606	5327	7402	6204	5216	9511	0145	4407
2225	4105	5575	5354	9190	9667	3896	3610	4398	8622	9613	8722	7660	8141	8922
1507	6559	4651	7610	9162	4502	0623	8353	5306	7346	5421	4992	6490	0868	7323
0525	7467	5629	1470	7150	7088	2736	4571	3323	5504	3615	8199	0720	6842	1583
3757	9743	8240	3837	1403	9785	0110	4526	6744	1897	7339	2223	2982	0299	4867
3934	6211	4903	0863	5501	7117	0980	9984	9837	7574	2885	6252	6631	9876	7689
8185	0935	0549	2719	0349	6359	8011	8187	0842	6450	5905	1492	0645	8788	4341
9698	8154	0394	8064	4653	0565	6530	8610	3923	5696	6513	2257	8723	5929	5121



## Disease Outbreak Questionnaire

Village Name		Date	
District Name		Vill No	

<b>Question 1</b>	
Has there ever been an outbreak of FMD in this village? (✓)	
Yes <input type="checkbox"/>	No <input type="checkbox"/>

**If Yes**

<b>Question 2</b>	
When was the last outbreak of FMD in the village? (Month and year that the first animal got sick.)	
Date	

<b>Question 3</b>	
At the time of the last outbreak, how many cattle and buffalo were there in the village?	
Pigs	
Cattle	
Buffalo	

**If No**

<b>Question 2</b>	
What is the earliest date since which you are sure there has been no FMD in the village?	
Date	

<b>Question 3</b>	
At that time, how many cattle and buffalo were there in the village?	
Pigs	
Cattle	
Buffalo	





# RGCS Random Point Sheet (Sheet 2)

Point ID	Visited	Number of villages
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		

# Appendix E

## Contents of CD

Survey Toolbox Win95 Installation Version

Survey Toolbox Win95 Runtime Version

Survey Toolbox DOS Runtime

Acrobat copy of text

EpiInfo 2000 Win95

EpiInfo 6.04 DOS

Adobe Acrobat (Install and RunTime)

# Index

- active surveillance, 3, 17
- animal health information, 3, 14, 84, 113
  - need for, 14
- animal restraint
  - cattle and buffalo, 87
  - goats, 94
  - horses, 93
  - Pigs, 92
  - sheep, 94
- anticoagulant, 101
- attitude, 112
  
- backing up, 132
- bias, 21, 24
  - measurement, 21
  - selection, 21
- bleeding pole, 89
- blood collection, 96
  - cattle and buffalo, 97
  - caudal vein, 97
  - chickens, 99
  - goat, 98
  - guidelines, 96
  - horses, 98
  - jugular vein, 97
  - pigs, 98
  - sheep, 98
- blood samples, 95
  - transporting, 100
- brainstorming, 217
  
- CapRecap, 187
- capture - recapture, 183
  - analysis, 187
  - background, 183
  - data collection, 186
  - data management, 186
  - data sources, 185
  - matching, 186
- censoring, 121
- census, 18
- centrifuge, 101
- children, 105
- clinical prevalence, 31
- clustering, 199
- Compare Prevalence, 167
- competitions, 216
- computer programs, 4
- computers, 124
- confidence interval, 23
- confidence level, 156
- cooperation, 111
- crush, 87
  
- data, 124, 126
- accuracy, 128
- categorical, 126
- checks, 129
- coding, 128
- completeness, 128
- continuous, 126
- conversion, 140
- dates, 126
- dichotomous, 126
- dictionary, 128
- entry, 131
- error checking, 133
- integers, 126
- linking, 139
- management, 175
- missing, 129
- Nominal, 126
- numbers, 126
- ordinal, 126
- processing, 127
- real numbers, 126
- saving, 132
- storage, 127
- text, 126
- types, 126
- yes/no, 126
- data analysis, 124, 141, 160
- database, 127
- dBASE, 140
- diagnostic tests, 32
- disease
  - impact of, 2
  - measures of, 26
  - patterns of, 2
- disease control, 2
- disease names, 109
- disease priorities, 116
- double entry system, 132
  
- encouraging participation, 106
- Epi Info, 4, 43, 51, 125, 129, 176
  - Analysis, 133
  - counting records, 134
  - creating a table, 130
  - data entry, 132
  - frequency tables, 134
  - histogram, 136
  - means, 135
  - recoding, 139
  - relate, 139
  - scatter, 138
  - select, 137
  - setting up data checks, 131
  - tables, 137
- error

- random, 24
- systematic, 21, 24
- Type I, 195
- Type II, 195
- estimation, 23
- evacuated tubes, 95
  
- faecal sample, 95
- field, 127, 129
- field trips, 218
- field work, 219
- FreeCalc, 196
  - analysis of results, 204
  - formulae, 206
  - Infinite Population Size, 207
  - Maximum Sample Size, 207
  - sample size, 196, 201
- freedom from disease, 190
  - data analysis, 204
  - herd or village, 194
  - Large-area surveys, 199
  - proving, 191
  - sample size, 194, 200
- frequency tables, 134
- fun, 113
  
- games, 216
- geographical information system, 67
- georeference, 73
- GIS, 67
- global positioning system , 67
- GPS, 67
- group discussions, 216
  
- halter, 88, 93
  - nose, 88
  - rope, 88
- hardware, 124
- Hazard Ratio, 182
- histograms, 136
  
- import, 140
- Incidence rate, 28
  - analysis of two data sources, 183
  - village-level, 170
- inference, 20
- information sources, 84
  - collecting specimens, 85
  - examining animals, 85
  - existing records, 84
  - interviews, 84
  
- landmarks, 120
- language, 109
- learning land marks, 214
- lesson planning, 213
- listening, 106
- Logrank test, 182
  
- maximum acceptable prevalence, 192
- measures of disease, 26
- median survival time, 180
- minimum expected prevalence, 192
- monitoring, 3
  
- nose grip, 92
  
- OIE, 3
- outbreak, 118
- outbreak history, 118
  
- passive surveillance, 3, 14
- payment, 112
- physical randomisation, 40
- pipette, 101
- plasma, 100
- population, 18, 195
- population size, 155
- practical activities, 219
- precision, 24, 156
- Prevalence, 26, 155
  - apparent, 34
  - Clinical, 31
  - comparison, 167
  - seroprevalence, 31
  - true, 34, 165
- Prevalence Analysis, 160
  - data inputs, 160
- Prevalence surveys, 150
  - conducting, 152
  - design, 152
  - first-stage sampling, 159
  - PPS, 153
  - probability proportional to size, 153
  - random geographic coordinate sampling, 154
  - RGCS, 154
  - sample size, 154, 156
  - second-stage sampling, 160
  - simple random sampling, 153
  - SRS, 153
- proportional allocation, 158
- proportions, 16
- prospective study, 170
  
- questions, 110
  
- race, 87
- Random Animal, 58, 160
- random error, 24
- random number tables, 42
- random numbers, 40, 41
  - selecting, 41
  - selecting with a computer, 43
- random sampling, 22
- Random Village, 50, 159, 175, 186, 198, 204
- ranking, 116, 217

- Rates, 16
- recoding, 138
- record, 127
- recording data, 86
- relative error, 156
- remote sensing, 72, 73
- replacement, 44
- representative samples, 20
- resolution, 73
- restraint, 87
- retrospective disease outbreak surveys, 171
  - activities, 172
  - complex analysis, 179
  - data analysis, 177
  - interview, 175
  - prerequisites, 171
  - sample size, 173
- retrospective study, 170
- RGCS, 68, 154
  - field procedures, 74
  - for ArcView GIS v.3, 69
  - for Windows 95, 68
- role playing, 218
  
- sample, 18
- sample size, 23, 154, 173
  - minimum cost, 201
- sampling, 38
  - convenience, 38
  - haphazard, 38
  - non-probability, 38
  - PPS, 48
  - probability, 38, 39
  - probability proportional to size, 48, 153
  - purposive, 38
  - random, 39
  - random geographic coordinate, 65
  - RGCS, 65
  - simple random, 153
  - stratified, 45
  - systematic, 44
  - techniques, 40
  - two-stage, 64
  - with replacement, 44
  - without a sampling frame, 65
  - without replacement, 44
- sampling frame, 115, 150
- sampling frames, 49
  - building, 55
  - selecting a sample from, 50
  - sources, 50
- sampling interval, 45
- seasonal differences, 178
- selecting animals, 54
  - identifying selected animals, 60
  - random number table, 56
  - with a computer, 58
- selecting subgroups, 137
  - selecting villages, 52
  - selection radius, 71
  - sensitivity, 33, 194
- Seroprevalence, 31
- serum, 101
- snare, 92
- social status, 108
- software, 4, 125
  - installing, 6
  - Requirements, 5
  - running, 6
- specificity, 33, 194
- specimen collection, 95
- stratification, 45, 158
- student-centred learning, 214
- surveillance, 3
  - active, 3
  - passive, 3
- survey
  - objectives, 111
- survey design, 152
- Survey Toolbox, 4
  - how to use, 4
- surveys, 18
  - costs, 154
  - Incidence rate, 170
  - retrospective disease outbreak, 171
  - to demonstrate freedom from disease, 190
  - which to use, 34
- Survive Size, 173
- syringe, 95
- systematic error, 21
  
- table, 129
  - creation, 129
- tables, 137
- teaching techniques, 214
- Tests, 32
  - combining, 205
  - in parallel, 205
  - in series, 205
- trainers
  - who should be a trainer, 212
- training
  - activities, 213
  - courses, 222
  - skills, 212
- True Prevalence, 165
- twitch, 93
- two-stage sampling, 64, 150
  
- under-reporting, 15
- uninterruptable power supply, 132
- units of interest, 18
- UPS, 132
  
- vacuum tubes, 95
- variance, 155

## 330 Index

- village calendar, 120
- village history, 120
- village interviews, 55, 104
  - activities, 114
  - introduction, 114
  - leaders, 106
  - order, 114
  - organising, 105
  - outputs, 114
  - sampling frame, 115
  - who should attend, 104
  
- warmers, 215
- weighting, 65
- women, 104, 108