



RURAL INDUSTRIES RESEARCH
& DEVELOPMENT CORPORATION

Epidemiology Workshop for Equine Research Workers

**University of Sydney,
Veterinary Conference Centre**

April 16th - 17th 1998

**Edited by Professor Reuben Rose
and Melissa Offord**

July 1998

RIRDC Publication No 98/61

RIRDC Project No WS978-09

© 1998 Rural Industries Research and Development Corporation.

All rights reserved.

ISBN 0 642 54078 0

ISSN 1440-6845

"Epidemiology Workshop for Equine Research Workers"

The views expressed and the conclusions reached in this publication are those of the author and not necessarily those of persons consulted. RIRDC shall not be responsible in any way whatsoever to any person who relies in whole or in part on the contents of this report.

This publication is copyright. However, RIRDC encourages wide dissemination of its research, providing the Corporation is clearly acknowledged. For any other enquiries concerning reproduction, contact the Communications Manager on phone 02 6272 3186.

Researcher Contact Details

Prof Reuben Rose
PMB 4
410 Werombi Road

Phone (02) 9351 2441
Fax (02) 9660 1548
Mobile (0419) 565 014 ,
Email: dean@doolittle.vetsci.su.oz.au

RIRDC Contact Details

Rural Industries Research and Development Corporation
Level 1, AMA House
42 Macquarie Street
BARTON ACT 2600

PO Box 4776
KINGSTON ACT 2604

Phone: 02 6272 4539
Fax: 02 6272 5877
Email: rirdc@netinfo.com.au
Internet: <http://www.rirdc.gov.au>

Published in July 1998

Printed on environmentally friendly paper by the DPIE Copy Centre

FOREWORD

The study of diseases in populations has long been the province of veterinarians dealing with herd and flock problems.

The use of epidemiological techniques have been used less in dealing with equine diseases and the majority of studies reported have involved analysis of risk factors for racetrack injuries.

This workshop on equine epidemiology is an initiative of the Equine Research and Development Advisory Committee of RIRDC to increase awareness of, and skills in epidemiology of equine diseases.

We have been fortunate in having two of the world's leading epidemiologists, Professors Roger Morris and Stuart Reid, being prepared to contribute to this workshop and provide an overview of epidemiological methods relevant to the horse industry.

The task ahead is to decide on priority areas where epidemiological techniques can be used to provide information that will benefit the different sectors of the horse industry.

I appreciate the fact that so many research workers involved in different areas of equine research have been prepared to commit two days to this workshop and feel sure that we will achieve outcomes that will be of benefit to the horse.

The workshop was part of the Corporation's equine research program which aims to assist in developing the Australian horse industry and enhancing its export potential.

Peter Core

Managing Director

Rural Industries Research and Development Corporation

LIST OF CONTRIBUTORS

- Stuart Reid
Veterinary Informatics and Epidemiology Group
Department of Veterinary Clinical Studies
University of Glasgow Veterinary School
and
Department of Statistics and Modelling Science
University of Strathclyde
Bearsden Rd., Bearsden, Glasgow G61 1QH
SCOTLAND
- Ian Robertson
Division of Veterinary and Biomedical Sciences
Murdoch University
Murdoch WA 6150
- Roger Morris
Massey University EpiCentre
Institute of Veterinary, Animal and Biomedical Sciences
Massey University
Palmerston North, New Zealand
- Chris Baldock
AusVet Animal Health Services
12 Thalia Court
Corinda QLD 4075
- Craig Bailey
Department of Veterinary Clinical Sciences
University Veterinary Centre (Camden)
410 Werombi Rd
Camden NSW 2570
- Simon More
School of Veterinary Science and Animal Production
The University of Queensland
PO Box 125
Kenmore QLD 4069
- Reuben Rose
Research Manager, Equine R&D Program
Rural Industries Research and Development Corporation
University Veterinary Centre (Camden)
410 Werombi Rd
Camden NSW 2570
- Jackie Lublin
10 King St
Balmain NSW 2041
- Nigel Perkins
Massey University EpiCentre
Institute of Veterinary, Animal and Biomedical Sciences
Massey University
Palmerston North, New Zealand
- Kathryn Knox
Veterinary Informatics and Epidemiology Group
Department of Veterinary Clinical Studies
University of Glasgow Veterinary School
Bearsden Rd., Bearsden, Glasgow G61 1QH
SCOTLAND

TABLE OF CONTENTS

Foreword	iii
List Of Contributors	iv
Why Epidemiology: An Introduction	1
Introduction.....	1
Outbreak Investigation; Where It All Started.....	1
The Diversity Of The Epidemiological Approach	2
What Will We Do?	2
Convincing The Clinicians	5
Summary.....	5
References	6
Some Measures Of Disease Occurrence And Effect	7
Introduction.....	7
Ratios, Proportions And Rates	7
Prevalence And Incidence.....	7
Stratification: Crude And Specific Rates.....	9
Relative Risk And Odds Ratio	9
Epidemiological Study Designs	11
Design And Analysis Of Observational Studies	12
Case Studies	12
Cross-Sectional Studies	13
Case-Control Studies	15
Cohort Studies (Prospective Or Longitudinal Studies).....	17
Analysing Data From Observational Studies	18
Odds Ratios.....	19
Attributable Risk	19
Attributable Fraction	20
Calculation Of Relative Risk, Attributable Risk And Attributable Fraction.....	20
Multivariate Analysis.....	21
References	22
How To Design A Successful Equine Epidemiological Research Study	23
Introduction.....	23
1. Defining The Research Question	23
2. Epidemiological Approach	24
3. Subjects.....	25
4. Variables	25
5. Data Gathering And Management.....	26
6. Statistical Issues In Study Management.....	26
Conclusion.....	33
References.....	34
Equine Morbillivirus: An Epidemiological Perspective	35
Summary.....	35
Introduction.....	35
The Mackay Outbreak	36
The Brisbane Outbreak	37
Similarities Between The Mackay And Brisbane Outbreaks	38
Cause Of The Outbreaks In Horses	38
Natural History And Transmission In Horses.....	39
Host Range	39
Fruit Bats - A Natural Wildlife Reservoir	39
Discussion.....	41
References	42
Wastage In The Racing Industry - Approaches To Study	43
Introduction.....	43
Identification Of Risk Factors For Musculoskeletal Racing Injuries	43
Use Of Survival Analysis For The Study Of Racing Careers	46
A Longitudinal Study On Injuries And Disease In 2- And 3-Year Old Thoroughbreds In Training	47
References	51

A Longitudinal Study Of Australian Racing Thoroughbreds: Performance During The First Years Of Racing	53
Summary	53
Introduction.....	53
Materials And Methods.....	54
Results	56
Discussion	66
Acknowledgments	68
References	69
Application Of Epidemiological Techniques To Studies Of Equine Disease	71
Introduction - An Overall Approach	71
Case Series Study.....	71
Case-Control Study	72
Cross-Sectional Studies	72
Cohort Study	73
Longitudinal Population Study	73
Intervention Study	74
Disease Process Studies.....	74
Modelling And Prediction.....	75
Synthesis Of An Approach	75
References	76
Multivariable And Multifactor Techniques: An Introduction	78
Introduction.....	78
Multivariable Logistic Regression.....	78
Analysis Of Variance	81
Making The Most Of Large Clinical Datasets	86
Introduction.....	86
Veterinary Clinical Data	86
Approach To Large Clinical Datasets	86
Investigation Of Guvs Hospital Database.....	88
Sensitivity, Specificity And Predictive Values	92
Introduction.....	92
The Perfect Test.....	92
Sensitivity And Specificity	93
Other Issues	94
Predictive Values.....	94
Examples.....	95
Serial Versus Parallel Testing	96
Trust.....	96
Some Final Comments	96
References	97
Clinical Trials: Design And Assessment	98
Introduction.....	98
Protocols	99
Experimental Unit	100
Randomisation	100
Trial Design	100
Design Sensitivity And Validity	101
Problems With Clinical Trials.....	101
Some Common Experimental Designs.....	103
 Appendix I:	 The RIRDC Equine Industry Programme: The Role for Epidemiological Research
 Appendix II:	 Workshop on Major Problems Facing the Equine Population in Australia: Prioritising Equine R&D Issues

WHY EPIDEMIOLOGY: AN INTRODUCTION

Stuart W. J. Reid

INTRODUCTION

Lord Kelvin, a native of Glasgow, Scotland and most famous for absolute zero, is credited with the statement “I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it: but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.” A useful maxim for any quantitative scientist but of particular relevance to the epidemiologist.

Epidemiology is the study of disease in specified populations and the quantitative assessment of the factors that determine its occurrence, distribution and severity. *Classical* epidemiology is primarily concerned with the statistical relationships between disease agents, both infectious and non-infectious; *ecological* epidemiology studies and describes (often mathematically) the ecological interactions between populations of hosts and infectious agents. Other sub-categories, e.g. molecular epidemiology, clinical epidemiology or environmental epidemiology, relate to the techniques and domains in which the quantitative tools are being applied. Often the techniques and approaches will be different, but the two ubiquitous components are a *population-based approach* and *quantification*.

Epidemiology is perhaps the basic science of clinical medicine. Advances in the molecular sciences have led to improved methods for detecting, treating and preventing disease but only relatively recently has the need to evaluate the efficacy of diagnostic, therapeutic and prophylactic intervention become apparent. Medical research now acknowledges that risk assessment and quantitative appraisal are pivotal in the implementation of advanced health care systems (Hiatt, and Goldman, 1994). Epidemiology, biostatistics, decision analysis and health economics are the “evaluative biological sciences” necessary for such investigations. With particular reference to the equine industry, in order to maximise the benefits to health and welfare of the thoroughbred by the most efficient means, quantitative assessment of available data from currently employed strategies must be performed. Whilst animal welfare is paramount, intervention at all levels must be carried out in the most cost effective manner, if the industry is to prosper and collaborative links with practitioners and trainers, in projects initiated by the industry, can ensure significance of findings to the field situation.

OUTBREAK INVESTIGATION; WHERE IT ALL STARTED

The application of epidemiology to epidemic investigation is well established:

- a) What is the case definition?
- b) What agent or agents caused the disease and what are the characteristics of the agent?
- c) What is the source of the agent?
- d) Which animals are susceptible?
- e) How is the disease transferred from infected to susceptible individuals?
- f) How can we best reduce risk, reduce disease and prevent infection?

But epidemiology is much more than this and over the next two days a selection of the different applications of the discipline will be presented.

THE DIVERSITY OF THE EPIDEMIOLOGICAL APPROACH

Some subdivide the subject by type of study

Experimental epidemiology

- Clinical trials
- Field trials

Observational epidemiology

- Follow up studies
- Case control studies

But we have also to consider other activities such as

- Modelling
- Outbreak investigation
- Surveillance
- Survey

Others prefer to use terms like

- Clinical epidemiology; clinical decision making, diagnostic tests
- Molecular epidemiology; molecular markers in population studies
- Genetic epidemiology; family and sibling studies
- Chronic disease epidemiology
- Pharmacological epidemiology

WHAT WILL WE DO?

Regardless of definition, most of the study types or disciplines in which epidemiology is applied have common activities. The epidemiologist will;

identify
describe
quantify

In addition the epidemiologist will;

hypothesise
compare
measure effect

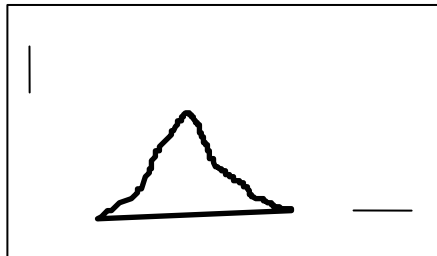
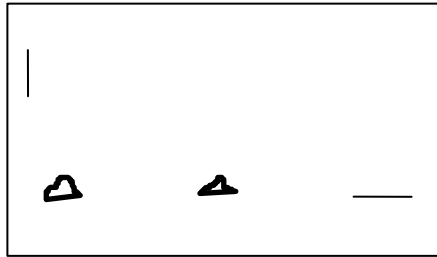
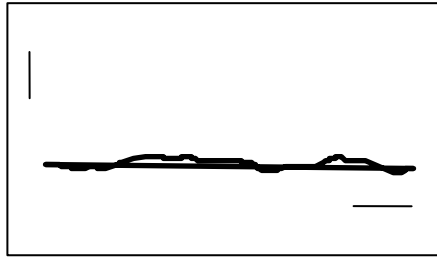
And may also;

model to describe
model to predict
intervene

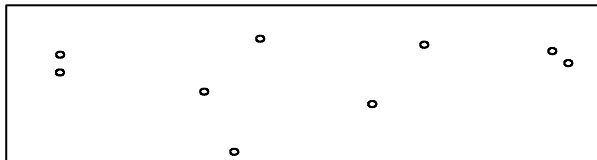
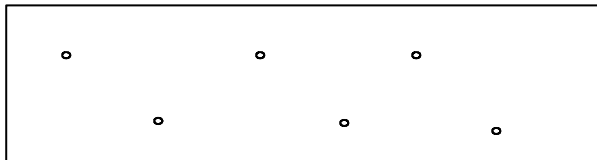
Key for the epidemiologist, then, is the need to be quantitative, adopt a population based approach and pay due regard to the interaction of animal, agent and environment, and the variability that occurs in such a complex relationship (See Thrusfield, 1995).

However, most of us come to the discipline for the first time in order to investigate a disease outbreak and similar to the structured approach we adhere to when examining an animal, the application or epidemiological techniques should follow a logical plan. It is now 20 years since Schwabe, Riemann and Franti (1977) exhorted the investigator to answer one question at a time and offered an example from which the approach is still generally applicable and allows us to consider some important issues.

1. Plot the epidemic curve for the disease outbreak; number of new cases versus time
2. Describe the shape of the curve; is it (a)endemic, (b)sporadic or (c)epidemic



3. Identify and characterise the population at risk of disease.
4. Describe the disease rate. Are there comparable groups of animals, neighbouring farms with which comparisons could be drawn?
5. Consider the spatial distribution of the disease.



6. Considering attack rates and overall prevalence, is there a chance that the disease could be infectious?
7. Is there an apparent latent period?
8. What are the biologically plausible exposure factors that predated the appearance of

clinical signs.

9. Does it appear as if the disease is transmitted directly or indirectly?

10. Is there evidence of vertical transmission?

It is worth remembering that many of the answers to these questions may rely on recourse to sources of information far removed from the diseased or dead animal and the epidemiologist may have to access data gathered by other workers. Such sources include:

- Government agencies
- Slaughterhouse information
- Referral centres
- Laboratory sources
- Company records
- Breed societies
- Private industry

For each of these sources to be of use to the investigator, there must be an understanding that the correct data have been recorded and that the quality of the data are adequate. It does not matter how much information is made available, if it is of poor quality the inferences from the data will be flawed.

The intellectual tools we need to perform the task in hand is;

- An understanding of the multifactorial nature of disease
- A commitment to measure
- Basic mathematical and statistical skills
- Humility

Epidemiology is a very catholic discipline. From the initial identification of the diseased animal through diagnostic work up, clinical and pathological examination and testing, tracing of contacts, identification of source, assessing economic impact and predicting future outbreaks it is unlikely that one individual or even one discipline will be exclusively involved. Epidemiology perhaps is the underlying or unifying science that brings together the experts from all of the appropriate fields.

There are some obvious advantages to observational epidemiology;

- **Non-invasive:** Projects can make use of existing information or study the incidence and distribution of naturally occurring cases of disease. No animal experimentation need be involved.
- **Evaluative:** Assessment of efficacy, efficiency and cost benefits of management techniques, therapy and current knowledge
- **Multidisciplinary:** The approach builds on strengths and brings together many interested parties. Projects can often be initiated by the equine industry or be conducted in collaboration with equine practitioners.

CONVINCING THE CLINICIANS

In many instances a sound epidemiological study will hinge on co-operation and understanding from clinical colleagues. And yet the clinician is perhaps the person for whom the epidemiological study is an every day event, dealing with the likelihoods associated with diagnosis, prognosis and treatment. The probabilities that have to be considered at some time during the investigation of the individual animal include:

- Chance of a horse being ill
- Chance of a horse having disease x
- Chance of horse having disease x after taking history
- Chance of horse having disease x after physical exam
- Chance of horse having disease x after diagnostic test 1
- Chance of horse having disease x after diagnostic test 2
- Chance of horse having disease x after post mortem
- Chance of horse having disease x after examining other animals

These will to a conclusion that may not be definitive but will be the most likely given the available evidence. In the following presentations the route to attaching figures to these largely subjective probabilities will be described; type of data required, how to collect the data, the usefulness of ancillary testing, and some real world examples are a few of the issues that we will address.

SUMMARY

The sessions ahead provide the opportunity to address the issues with the specific focus on equine disease. The studies which be presented demonstrate that, for the horse, epidemiology is still in its youth. What is the size of our population at risk. What caused the sudden deaths of trainer and horses in Queensland? What is the risk associated with racing at particular race-tracks? How do we make use of data that already exist? What do my clinical pathology results really mean? What respiratory pathogens are really associated with disease? In each case an epidemiological approach is appropriate.

And so back to Kelvin. Besides his many no, contributions to science, Kelvin is also alleged to have said that heavier-than-air flying-machines were an impossibility! Give or take two standard deviations, of course.

REFERENCES

- Hiatt, H. and Goldman, L., (1994) Making Medicine More Scientific. *Nature* **371**, 100.
- Schwabe, C.W., Riemann, H.P. and Franti, C.E. (1977) In: *Epidemiology in Veterinary Practice*. Lea and Febiger, Philadelphia, 303pp.
- Thrusfield, M.V., (1995) In: *Veterinary Epidemiology*. Blackwell Science Ltd, Oxford.

SOME MEASURES OF DISEASE OCCURRENCE AND EFFECT

Stuart Reid

INTRODUCTION

When charting an unknown or new environment, and in common with foreign travel, it pays dividends to have an appreciation of the language and the currency at an early stage. In this section we will concentrate on the terminology associated with measures of disease frequency and the units in which they are expressed.

RATIOS, PROPORTIONS AND RATES

1. Ratios

A ratio is composed of a number on top (numerator) and a number on bottom (denominator) which are mutually exclusive frequencies.

$$a / b$$

where a is not included in b . The ratio is often expressed by dividing the smaller value into the bigger value.

For example,

- the ratio of geldings to fillies in a yard may be 20/25. This can be expressed as 20:25 or 1:1.25.
- the ratio of diseased to healthy animals in a disease outbreak, 4/60, or 1:15.

Note that in both of these examples the two frequencies *must* be independent.

1. Proportions

A proportion is a fraction where a is included in b , (*cf* ratio). As such the proportion is dimensionless and is bounded by 0 and 1. This measure is often converted to a percentage by multiplying by 100. Proportions are frequently used where the numerator is the frequency of diseased animals and the denominator is the population.

Using the same examples as above;

- the proportion of fillies in the yard is 25/45 or 0.56.
- the proportion of diseased animals in the population is 4/64 or 0.063.

1. Rates

Rates may be thought of as proportions in which there is consideration given to an event in relation to time. In this case the frequency of the event during a specified time is in the numerator, and the denominator is the "population time". So

$$\frac{a}{b}$$

Note again that a is included in b .

PREVALENCE AND INCIDENCE

In epidemiology we may be interested in assessing a situation at one point in time or we may wish to describe events over a specific period of time. The measures appropriate for the different approaches are prevalence, incidence and cumulative incidence.

1. Prevalence

Also referred to as point prevalence, this is a proportion and is the number of animals affected at time t , divided by the total number of animals at risk at time t .

$$\text{Prevalence} = \frac{\text{Number of cases at time } t}{\text{Total number of individuals at risk at time } t}$$

Called to an outbreak of acute larval cyathostomosis in a riding stables, it is observed that 6 of 22 horses are diarrhoeic. The prevalence is therefore $6/22$ or $.272$ or 27.2% .

2. Incidence rate (or Incidence density rate)

This is a measure of the speed at which disease is occurring and is defined as the number of new cases divided by the total amount of time at risk from disease for all animals. An animal is no longer at risk if it leaves the population or becomes diseased. In essence the denominator is the total animal-disease-free-time. The incidence rate can range from 0 to infinity.

$$\text{Incidence} = \frac{\text{Number of new cases during period}}{\text{Total animal - time at risk during same period}}$$

The choice of units for the denominator will depend on the disease being studied. For example;

- 2 cases of neonatal septicaemia per foal-week
 - 20 cases of ocular squamous cell carcinoma per 1000 horse-years
- In this latter example the 20 cases could be derived from 100 horses studied for 10 year or 40 horses for 25 years.

The mortality rate is a special kind of incidence rate.

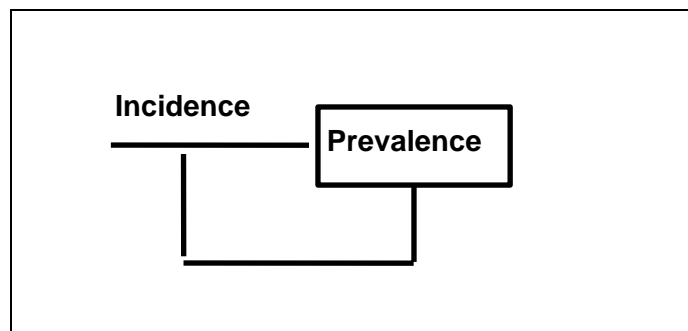
Prevalence Versus Incidence

Note that the duration of the disease can have a profound effect on the prevalence, as can the mortality associated with the disease. Diseases with high mortality rates have low prevalences. The prevalence is therefore a function of the disease duration and the incidence, where incidence can be thought of as the flow or rate at which new cases occur.

$$\text{Prevalence} = \text{Mean Duration} \times \text{Incidence}$$

A consequence of this is that for chronic disease such as certain types of neoplastic disease, prevalence measures are preferred, whilst for diseases with a high rate of new cases, such as infectious diseases, incidence measures are preferred.

Incidence is a function of the number of susceptible animals which in turn is related to prevalence, the number of diseased animals at any one time, which, when increasing means that although there may be potentially more infectious animals, the number of susceptibles is reduced.



So with both the number of susceptibles and the number of diseased animals influencing the rate of appearance of new cases, together with characteristics of the agent, it is clear that the situation is complex. In fact it is the combination of these factors that give rise to the bell shaped epidemic curve of a disease outbreak with which we are familiar.

1. Cumulative Incidence Rate

The Cumulative Incidence Rate (CIR) includes the concepts of “animal-time at risk” and a specified period of time but it should be remembered that although it is a proportion it is really regarded as rate. It is defined as the number of new cases that arise during a specified period of time as a proportion of a fixed population that is at risk of disease during that period.

$$\text{Cumulative Incidence Rate} = \frac{\text{Number of new cases during period}}{\text{Total number of animals at risk during same period}}$$

When using CIR it is essential that the period of time be specified and the CIR can be regarded as the average risk for the population during that period. It is also worth noting that the longer the period, the greater the CIR emphasising the need to define the duration of the period.

The CIR is one of the more widely used measures and important examples include;

- The attack rate (often referred to as morbidity)
- The case–fatality rate

STRATIFICATION: CRUDE AND SPECIFIC RATES

Sometimes we may wish to calculate the measures we have described above for a whole population regardless of important animal attributes that we suspect might be important in determining disease occurrence, such as age or gender. In the case of a single summary measure we generally refer to this as a *crude* measure. If we subdivide or stratify the population we refer to it as a *specific* measure. For example the *gender specific* incidence of pituitary adenoma is higher in mares than in geldings; the *age-specific* prevalence of alimentary lymphosarcoma is highest in young horses.

The concept of stratification leads us to consider the fact that we may have two groups between which we wish to compare disease occurrence. As we shall see the type of epidemiological study that we are performing dictates the best measures and therefore the best methods of comparison. However, there are some measures of effect that are commonly encountered and we shall consider here the two most important

RELATIVE RISK AND ODDS RATIO

In order to appreciate the methods of comparison for measuring an effect, it is easiest to think it terms of risk: How many times more at risk of disease are horses exposed to procedure A relative to horses that are not exposed to procedure A?

This is given by the ratio of the IRs or CIRs. Consider the table:

	Disease status	
	Positive	Negative
Exposed	a	b
Unexposed	c	d
Relative risk	=	$(a / a + b) / (c / c + d)$

Note that to have calculated the CIR or IR we require to know the size of the population at risk for exposed and unexposed. Suppose in another situation we do not know the total size of the population at risk but we do have cases and non-cases that have been exposed or not exposed to a factor that we have chosen to investigate retrospectively.

	Disease status	
	Positive	Negative
Exposed	a	b
Unexposed	c	d

This time we look at the ratio of the odds of exposure in cases to the odds of exposure in non-cases:

$$\begin{aligned}
 \text{Odds ratio} &= [(a / a + c) / (c / a + c)] / [(b / b + d) / (d / b + d)] \\
 &= (a / c) / (b / d) \\
 &= (ad / bc)
 \end{aligned}$$

Now if we consider the case where the disease is rare and

$$\text{Relative risk} = (a / a + b) / (c / c + d)$$

It follows that a will be much less than b and c will be much less than d .

Therefore,

$$\begin{aligned}
 \text{Relative risk} &= (a / b) / (c / d) \\
 &= (ad / bc) \\
 &= \text{Odds ratio}
 \end{aligned}$$

So we can conclude that in certain circumstances, what we are calling the Odds Ratio will approximate to the Relative Risk. However for a fuller understanding of measures of effect we must consider the different types of study design.

EPIDEMIOLOGICAL STUDY DESIGNS

Ian Robertson

The primary aim of most studies in veterinary science is to identify the cause of disease, or the factors that predispose to disease so that preventive measures can be implemented. The initial process in any disease investigation is to undertake a descriptive study which will answer specific questions such as: What is the prevalence of disease? When does the disease occur? Where is it present? What type of animals does it affect? Frequently these descriptive studies identify factors which may be involved in the occurrence of the disease. To determine whether these factors are associated with the disease, specific studies must be conducted. These epidemiological studies are frequently called *observational studies* because the veterinarian/researcher observes what is happening or has happened without intervening in the natural progression of disease events (Martin 1990a). For example, Tinker *et al* (1997) conducted an observational study investigating risk factors for colic in horses. They found that horses which had multiple changes of diet in a year were more likely to develop colic than were horses with a consistent diet. Knowing this, advice could then be given to owners to reduce the incidence of colic in their horses.

Observational studies involve making comparisons. Two main types of comparisons are performed:

1. the disease frequency in naturally exposed animals (factor +) is compared with the disease frequency in non-exposed individuals (factor -)
2. the frequency of a factor in naturally diseased animals (cases) is compared with the frequency of that factor in non-diseased individuals (controls).

Three types of study designs are used to perform these comparisons :- cross-sectional studies, case-control studies and cohort studies. A fourth type of study, the case study, is also an observational study but does not have a control group and is therefore of less value. A key for classification of observational studies, adapted from Smith (1995), is outlined in Table 1.

Factors investigated in observational studies can be:

- features of the animals
eg male vs female, stallion vs gelding, < 4 years of age vs \geq 4 years of age, Thoroughbred vs Standardbred, pregnant vs non-pregnant etc
- management/husbandry features
eg fed > 2.5 kg concentrate/day vs no concentrate, stabled vs not-stabled, Stakes race vs non-stakes race, barrier \geq 7 vs barrier < 7 etc

The information gained from observational studies is relatively cheap and easily obtained as the studies are directed towards the animal in its *natural environment*, rather than the artificial environment often found with experimental studies. Because the study is conducted in a natural setting, problems of extrapolating results back to the general population are usually reduced. Researchers can also test a much broader range of hypotheses, and therefore factors, than are usually feasible under experimental conditions. However a disadvantage with observational studies is that the groups under observation may differ in a wide range of characteristics. These characteristics may confound the findings of the study. Care must therefore be taken when selecting groups for observational studies, interpreting the findings and when extrapolating the results back to the general population. This disadvantage can be overcome by undertaking experimental studies or trials and these are often conducted after observational studies. Although experimental studies are more scientifically rigorous, observational studies are the only feasible method of studying many questions of risk (Smith, 1995).

DESIGN AND ANALYSIS OF OBSERVATIONAL STUDIES

In observational studies a series of steps are undertaken (Frankena and Thrusfield, 1997).

1. The objectives of the study are defined
2. The target population is described
3. The sampling method is selected and sample size calculated
4. Disease and exposure factors are measured in the sample
5. Bias (selection, misclassification, information or recall bias and confounding) is evaluated
6. Data is validated
7. Data is analysed
8. Findings are reported

If care is taken in the planning, implementation and analysis of observational studies, risk factors can be identified to allow preventive measures to be instigated.

CASE STUDIES

Case studies are the least useful of the observational studies and in fact many would argue they are not a true “epidemiological study”. They involve a detailed study of a small number of cases (sometimes up to 100) of disease. Their main value lies in the detailed study of affected animals eg Maxson and Beef (1997) investigated bacterial endocarditis in ten horses to evaluate the value of echocardiography as a diagnostic tool, while Hillyer and Mair (1997) investigated 58 cases of recurrent colic to determine the aetiology and clinical features of horses with this condition.

Case studies are often descriptive in nature and precede the “true” observational studies. They help to generate hypotheses and determine which factors should subsequently be investigated in an observational study. Case studies are very good for rare diseases and for new or emerging diseases. The disadvantage with them is that no control group is used and there is always the danger that the cases examined may not be representative of the normal disease situation.

In the study of Maxson and Beef (1997) on bacterial endocarditis, it was concluded that there was no breed or sex predilection, however without a comparison group such an observation should have not been made. This is one of the major limitations with case studies in that association of disease with specific factors can not be determined. The other three types of observational studies all have comparative or control groups, allowing the identification of risk factors.

Table 1 Key For Classification Of Study Designs

(adapted from Smith, 1995)

1a	Subjects being investigated experience experimentally induced disease, condition or intervention	Clinical Experiment
1b	Subjects under study experience naturally occurring disease, condition or intervention	Go to 2
2a	In-depth study of a few affected animals without a comparison group	Case study or report
2b	Comparison group present	Go to 3
3a	Cross sectional study: All observations on the animals are made at one point in time in the course of that individual's illness	Go to 4
3b	Longitudinal study: Subjects followed over a period of time	Go to 5
4a	Cases selected from a group of diseased animals; non-cases (controls) selected to resemble cases.	Case control study
4b	Animals are selected from a population and then examined to see if they have the disease and factor being studied	Cross-sectional study
5a	No intervention (natural study or observation)	Cohort study
5b	Intervention	Clinical trial

CROSS-SECTIONAL STUDIES

A cross-sectional study examines the association between disease and risk factors in a sample or cross-section of a population at one particular instant in time. *The sample is first selected (ideally randomly) and then the presence or absence of disease and risk factors are determined simultaneously.* A variety of methods for selecting the sample can be used eg simple random sampling, systematic sampling, stratified sampling, cluster or multistage sampling (see Martin 1990b).

In all observational studies, after the data is collected, a 2 x 2 contingency table is usually constructed (see Table 2). At the beginning of a cross-sectional study only n or $(a + b + c + d)$ or the sample population is selected. Then the animals are examined so that a , b , c and d can be filled in the table.

Table 2 A 2 x 2 contingency table for observational studies

	Disease		Total	Attack Rate
	Present	Absent		
Exposed (+)	a	b	$a + b$	$a \div (a + b)$
Risk Factor				
Unexposed (-)	c	d	$c + d$	$c \div (c + d)$
Total	$a + c$	$b + d$	$a + b + c + d = n$	

In cross-sectional studies the following can be calculated from Table 2:

proportion of population that are exposed to the factor or factor + ie $(a + b) \div n$

proportion of population with the disease $(a + c) \div n$

proportion of diseased animals in the exposed group $a \div (a + b)$

proportion of diseased animals in the non-exposed group $c \div (c + d)$

proportion of exposed animals in the diseased group $a \div (a + c)$

proportion of exposed animals in the non-diseased group $b \div (b + d)$

Therefore cross-sectional studies, as well as testing the hypothesis under study, can also be used to collect information about the population structure (exposed or factor + vs non-exposed or factor - eg the proportion of horses fed more than 2.5 kg of concentrate per day).

Cross-sectional surveys are usually employed as one of the first steps in the investigation of causal associations. The important features are :

- The study is conducted at one point in time.
- The animals are selected as a cross-section of the whole population
- The animals are not grouped in either disease or risk factor categories **before selection**
- The animals are grouped into disease and risk factor categories **after** all animals have been selected.
- Disease **prevalence** rather than incidence is recorded, however if a series of cross-sectional surveys are carried out on the same population inferences can be made about the incidence of disease.

One of the major problems associated with the design and analysis of cross-sectional surveys is avoiding confounding variables. It must be ensured that the sample of animals studied is truly representative of the total population and there are no other unidentifiable factors associated with the variables under study.

In cross-sectional studies the following statistics are calculated from Table 2:

Relative risk (RR) = $[a \div (a + b)] \div [c \div (c + d)]$
 = Attack rate in the exposed animals \div Attack rate in the non-exposed animals

Attributable risk (AR) = $[a \div (a + b)] - [c \div (c + d)]$
 = Attack rate in the exposed animals - Attack rate in the non-exposed animals

Attributable fraction = $AR \div [a \div (a + b)]$
 = $[RR - 1] \div RR$

These statistics will be explained later in this session.

Advantages of cross-sectional studies

1. Generally rapid and inexpensive
2. Can study several possible factors at one time
3. Gives direction for further studies
4. If the animals have been randomly selected, the data is more representative (disease prevalence and proportion of population with risk factors) of the general population than with case-control studies.

5. There is minimal risk to the subjects
6. The prevalence can be determined
7. The frequency of risk factors in the population can be calculated.

Disadvantages of cross-sectional studies

1. If disease is rare large numbers of animals are required
2. They cannot determine the outcome of disease control programs directed against the factor.
3. Difficult to determine cause and effect (Which came first?) as observations on the presence of disease and factors are made at the same time - ie can't make causal associations.
4. If effects are short-lived such as with a short lasting titre, cross-sectional studies are less effective.
5. Cannot estimate the incidence of disease

Examples of cross-sectional studies

Slater and Hood (1997) conducted a cross-sectional study of three groups of horses in Texas to investigate factors influencing hoof wall problems. They reported that 28% of horses had some type of hoof wall problem as identified by their owners ie they measured the prevalence of the condition. When they analysed their data they found that horses which had received oral hoof supplements were 4.8 times (relative risk) more likely to have hoof wall problems than non-supplemented horses. This was not unexpected as those horses with hoof problems were more likely to have been treated by their owners.

Atwill *et al* (1996) conducted a cross-sectional study of horses in New York state to identify environmental, host and management factors associated with an increased risk of seropositivity to *Ehrlichia risticii*. They found that the risk of seropositivity was associated with time spent in a stall, frequency of application of fly spray and Standardbreds were two times more likely to be seropositive than Thoroughbreds.

CASE-CONTROL STUDIES

A case-control study is one that starts with the identification of a group of animals with the disease (*cases* - $a + c$) and a group of animals without the disease (*controls* - $b + d$). These animals are then examined for the presence or absence of the factors under study. It is generally conducted as a *retrospective* study. In a case-control study only the numbers $a + c$ and $b + d$ are known at the start of the study.

Cases and controls are sometimes matched for possible risk factors not under investigation eg Cohen *et al* (1997) matched cases (Thoroughbreds that had musculoskeletal injuries during racing) with controls (two uninjured horses randomly selected from the same race as the injured horses) to overcome potential bias of the standard of race. In contrast Bailey *et al* (1997) randomly selected their controls from horses listed in the Australian Race Results to compare with cases of musculoskeletal injury reported by veterinarians at race courses. In some studies cases and controls can be groups of animals or farms eg Ross and Kaneene (1995) used a case-control study to study Eastern Equine Encephalomyelitis in USA where infected farms were used as cases rather than individual positive horses.

When selecting cases it is important that a clear definition of what constitutes a case and what doesn't is made to minimise bias. In some studies missing data can be a problem and Bailey *et al* (1997) overcame this problem by excluding all horses with any missing data from their study.

The important features of case-control studies are :

- The animals are selected because of the presence or absence of disease.
- The information on risk factors is usually obtained retrospectively from a data base, owners records or memory.
- Risk factors are not considered in the selection of animals.

To measure whether or not a factor is more common in the diseased group than in the control group the odds ratio (OR - discussed later) is calculated rather than the relative risk used for cross-sectional studies. In a case-control study **you cannot estimate the attack rate** as you have selected a biased population (not representative of the general population) ie ***the prevalence and incidence of disease can not be calculated*** and no information is collected about the frequency of the factors in the population.

Advantages of case-control studies

1. Usually retrospective, therefore uses pre-existing data
2. Quick and cheap to conduct.
3. With diseases of low prevalence, far fewer animals are used than with cross sectional or cohort studies.
4. Can study several possible factors at one time.
5. Gives direction for further studies
6. Good for diseases with long incubation periods
7. No risk to subjects

Disadvantages of case-control studies

1. Representativeness of data is difficult to guarantee: often NO knowledge of the bias due to death or culling etc. of diseased stock. May be relying on recall by owners (can't validate data - recall bias).
2. Difficulty arises in trying to determine the cause and effect relationship between the disease and the factor studied. It is difficult to know which came first.
3. Information may be incomplete
4. Cannot estimate prevalence of disease in the general population as the selection of cases and controls is not usually in the same proportion as in the population.
5. Cannot estimate the proportion of exposed (factor +) and non-exposed (factor -) animals in the population
6. Statistical associations discovered do not prove causal relationships.
7. They can not determine the outcome of disease control programs directed against the factor.
8. Selection of an appropriate comparison/control group may be difficult

Examples of case-control studies

Bailey *et al* (1997) conducted a case-control study to determine risk factors for musculoskeletal injury in racing Thoroughbreds. Horses at greater risk were older (horses older than four years of age were 1.8 times (odds ratio) more likely to be injured than younger horses), started from a wider barrier position (horses starting from a barrier position wider than 7 were 1.9 times more likely to be injured than horses starting from an inside barrier) and were in a stakes race (horses in a stakes race were 2.3 times more likely to suffer a musculoskeletal injury than those not in a Stakes race).

Reeves *et al* (1996) used a multi-center case-control study to identify risk factors for colic in horses. They found that Arabian horses were 2 times more likely to develop colic than were Thoroughbreds while Standardbreds were nearly half (OR 0.6) as likely to develop colic as Thoroughbreds. Horses that had access to outdoor enclosures without continuous drinking water were 2.2 times more likely to develop colic than horses in outside enclosures with an adequate supply of water. Although most case-control studies are retrospective ie existing data from records is used, Reeves *et al* (1996) conducted a prospective case-control study. At each hospital for every colic case that was identified, a control horse was selected at random from all non-colic admissions to that hospital during the week.

COHORT STUDIES (PROSPECTIVE OR LONGITUDINAL STUDIES)

A **cohort** is a term used to describe any group of animals which is followed or traced over a period of time. A cohort study is one in which a group with a specific factor (factor + or $a + b$ in Table 2) and a group without the factor (factor - or $c + d$ in Table 2) are followed for a period of time. The frequency of disease is then compared in the different groups. In a cohort study only the totals $a + b$ and $c + d$ are known at the start of the study. After the study is completed the individual cells (a , b , c and d) in Table 2 can be filled in.

The important features of cohort studies are:

- The groups of animals are followed through time.
- The animals are selected because of the level of exposure to the risk factor ie $a + b$ and $c + d$ (Table 2) are predetermined
- The risk factor can already be present in the population or it can be introduced by the observer (the classical experiment or clinical trial where animals are allocated to treatments at random is a specific type of cohort study).
- Disease is not considered in animal selection. The disease status is monitored over a period of time to determine the frequency of disease.

Usually the two cohorts are specifically sampled from the general population and therefore one does not gain information about the frequency of the factor or of the prevalence of disease in the general population. As cohort studies involve the collection of data over time, they provide information on the **incidence** of disease as opposed to prevalence in cross-sectional studies.

Advantages of cohort studies

1. The groups in the cohorts are defined in terms of characteristics prior to the appearance of disease.
2. Exposure to the disease factor can be more closely defined than with case-control studies.
3. The incidence of disease is measured in exposed and non-exposed groups
4. They provide useful information for determining the outcome of disease control programs.

Disadvantages of cohort studies

1. They may require long periods of time before results emerge, especially with diseases with a long incubation period. In animal populations life spans are shorter than in humans which reduces this problem (This can be overcome if retrospective cohort studies are used, however good records are required).
2. They are expensive.
3. They may require large numbers especially if the disease is rare.

4. Members of the cohorts may disappear, die, be sold or move away which can introduce bias.
5. Cannot estimate proportion of exposed/non-exposed in general population.

Examples of cohort studies

Most cohort studies are carried out prospectively ie the animals are divided into two groups based on their exposure and are followed over time to see what happens. Tinker *et al* (1997) conducted a prospective cohort study to investigate risk factors associated with colic. They used multivariate logistic regression to demonstrate that animals 2 - 10 years of age were 2.8 (95% confidence intervals: 1.2, 6.5) times more likely to develop colic than animals < 2 years, horses with a history of previous colic were 3.6 (95% CI 1.9, 6.8) times more likely to develop colic than horses with no history, and horses fed high levels of concentrate during the year (> 5 kg /day dry matter) were 6.3 (1.8, 22) times more likely to develop colic than horses not fed concentrate. In contrast feeding whole grain was protective (OR 0.4, 95% CI 0.2, 0.8) compared with not feeding whole grain. These findings demonstrated that diet and dietary changes were important risk factors for colic in horses.

Retrospective studies can be carried out *if records are accurate*. Storgaard Jorgensen *et al* (1997) conducted a retrospective cohort study to investigate the significance of radiographic findings with the racing performance of Standardbred trotters. From a group of horses that had been x-rayed as 1 - 1.5 year olds, they grouped the horses into those with nil or varying grades of radiographic abnormalities. Information on the racing performance was then collected from existing records. They found that there were no significant associations between the presence or type of radiological abnormalities and the subsequent performance and longevity of racing.

Gibson *et al* (1997) conducted a prospective cohort study comparing the development of superficial digital flexor tendonitis in Thoroughbreds which were managed non-surgically and Thoroughbreds which had a superior check desmotomy. They found that horses which underwent superior check desmotomy were 5.5 times more likely to develop suspensory desmitis than horses treated non-surgically. They concluded that superior check desmotomy did not appear to offer an advantage over non-surgical treatment in preventing recurrent or new injuries and horses with superior check desmotomy were at a greater risk of developing suspensory ligament injuries than horses managed non-surgically.

ANALYSING DATA FROM OBSERVATIONAL STUDIES

The results from observational studies are usually entered into a 2 x 2 contingency table (Table 2) however the tables can be larger eg if when there are more than 2 categories of risk factors such as with entire males, neutered males, entire females and neutered females.

Tables can be analysed using standard statistical tests for association (eg. chi square test), these however **give no indication of the biological importance** of the relationship. Three measures of RISK which show biological importance can easily be calculated - *Relative risk*, *odds ratio*, and *attributable risk*.

Relative risk is the ratio of the disease rate (attack rate) in individuals exposed to a risk factor divided by the rate in individuals that were not exposed to the factor. The relative risk indicates the chance of an event occurring in a study group (eg exposed or treated group) relative to the chance of the event occurring in a reference group (eg non-exposed or non-treated group).

$$\text{ie Relative risk} = [a \div (a + b)] \div [c \div (c + d)]$$

Relative risk is sometimes called risk ratio, incidence rate ratio or the prevalence ratio. It has no units as it is the proportion of two attack rates. It is an indication of the strength of association of the factor with the occurrence of the disease. The higher the relative risk the greater the association.

- If there is no association the relative risk will be close to 1
- If the factor has a relative risk < 1, it is called a sparing or protective factor ie the presence of that factor protects against disease.

The significance of the relative risk is determined by calculating 95% confidence intervals, if these confidence intervals include the value 1.0, the relative risk is considered not to be significant. Relative risk can be calculated in cross-sectional and cohort studies however not in case-control studies as we don't know the rate of disease in the exposed and non-exposed groups.

ODDS RATIOS

Where the population selected is biased and does not represent the true general population (eg. with many case records and with all case-control studies) another measure of risk is used instead of relative risk - the **odds ratio**.

$$\text{Odds ratio} = (a \times d) \div (b \times c)$$

The odds ratio approximates the relative risk when the attack rate is low (< 5%) as $a + b$ is almost equal to b and $c + d$ is almost equal to d . It deviates from the relative risk when the attack rate is high. The odds ratio is generally used when analysing case-control studies however *it can be used for all study designs* and is used in multivariate analysis such as logistic regression.

The odds ratio is interpreted in the same manner as relative risk. It measures the odds or chances of disease being present when a factor is present compared to the odds of disease being present when the factor is absent. The significance of the odds ratio is determined by calculating 95% confidence intervals, if these confidence intervals include the value 1.0, the odds ratio is considered not to be significant.

The 95% confidence interval can be calculated by:

$$e^{(OR \pm 1.96 * \sqrt{[(1 \div a) + (1 \div b) + (1 \div c) + (1 \div d)])}}$$

where $e = 2.718$, and \wedge is raised to the power of.

ATTRIBUTABLE RISK

Some disease is usually present in the non-exposed or factor negative group, therefore not all disease in the factor positive group can be due to that factor. To calculate how much disease can be attributed to a factor we calculate the **attributable risk** (AR). This is the difference between the attack rates in the exposed and non-exposed groups and has the same units as the initial rates. It determines the amount of disease that can be attributed to exposure to the factor. It gives an indication of the **biological importance** of the association and indicates the likely effect of a preventative campaign to remove exposure to the risk factor ie if there is a high AR, removal of the factor will have a significant influence on the level of the disease in the population, whilst if the AR is low, removal of the factor will not have a great influence on the level of disease. If a factor is protective it will have a negative attributable risk.

$$\begin{aligned} \text{Attributable risk} &= \text{attack rate exposed} - \text{attack rate not exposed} \\ &= [a \div (a + b)] - [c \div (c + d)] \end{aligned}$$

The attributable risk can not be calculated for case-control studies as we can not determine the attack rates.

ATTRIBUTABLE FRACTION

The attributable fraction (AF) is the attributable risk divided by the attack rate in the exposed group.

$$AF = AR \div [a \div (a + b)]$$

this is the same as

$$AF = (RR - 1) \div RR$$

Attributable fraction measures the proportion of disease that can be attributed to the factor being studied. eg if we have an AF of 70% it means that 70% of disease can be attributed to exposure to the factor which indicates that if we remove the factor the level of disease will drop by 70%. It therefore indicates the benefit in controlling exposure to the factor.

The previous formulae can only be used in cohort and cross sectional studies. However the AF can be estimated in case-control studies with the formula:

$$AF \approx (OR - 1) \div OR$$

Other epidemiological statistics such as population attributable risk and population attributable fraction can also be calculated to measure how much a risk factor contributes to the level of disease at the population level. A relatively weak risk factor that is quite prevalent could contribute more to disease incidence in a population than a stronger risk factor that is rarely present (see Frankena and Thrusfield, 1997).

CALCULATION OF RELATIVE RISK, ATTRIBUTABLE RISK AND ATTRIBUTABLE FRACTION

Doll and Hill (1962) studied cohorts of British Doctors and investigated the influence of smoking on lung cancer. The doctors were divided into 4 cohorts according to how many cigarettes they smoked each day, and after 11 years they obtained the results outlined in Table 3 (a cohort study).

Table 3 Calculation of relative risk, attributable risk and attributable fraction

Cigarettes per day (in 1951)	Annual death rate per 1000	Relative risk	Attributable risk	Attributable Fraction(%)
None	0.07	1	-	-
1-14	0.57	8.1	0.50	87.7
15-24	1.39	19.9	1.32	95.0
>25	2.27	32.4	2.20	96.9

Therefore, it can be interpreted that the risk of dying from lung cancer is 32.4 ($2.27 \div 0.07 =$ relative risk) times as great for someone who smokes more than 25 cigarettes per day compared to someone who does not smoke. The attributable risk of 2.2 ($2.27 - 0.07$) suggests that a campaign to stop smoking, if successful, could drop the death rate from lung cancer by 2.2 per 1000. Note that there is a background effect of 0.07 per 1000 who died of lung cancer even though they did not smoke. These cases are due to other carcinogens or environmental factors. The attributable fraction for people who smoke more than 25 cigarettes is 0.969 ie $2.2 \div 2.27$ [or $(32.4 - 1) \div 32.4$] which means that 96.9% of lung cancer in heavy smokers can be attributed to cigarette smoking.

MULTIVARIATE ANALYSIS

The statistics described above are easy to calculate if only one factor is considered, however usually many potential risk factors are under investigation in the same study. There are a number of multivariate analytical techniques available for dealing with a number of factors in the same analysis. These methods show which of the risk factors or combination of risk factors are most important in the development of the disease. Examples of these are:-

- Multiple regression
- Logistic regression
- Discriminant analysis
- Log-linear models etc.

Most of these are now available in computer statistical packages, however the appropriate analysis depends on the nature of the data and it is advised to consult with an epidemiologist/statistician when planning, conducting and analysing a project. The method of interpreting the findings is the same as discussed earlier eg Bailey *et al* (1997) used logistic regression to show that horses at greater risk of musculoskeletal injury were older (horses older than four years of age were 1.8 times - odds ratios OR 95% CI 1.0, 3.0 more likely to be injured than younger horses), started from a wider barrier position (horses starting from a barrier position wider than 7 were 1.9 times - OR 95% CI 1.1, 3.2 more likely to be injured than horses starting from an inside barrier), or were in a stakes race (horses in a stakes race were 2.3 times - OR 95% CI 1.1, 4.6 more likely to suffer a musculoskeletal injury than those not in a Stakes race). Multivariate analysis will be discussed in more detail in a later session.

REFERENCES

- Atwill ER, Mohammed HO, Lopez JW, McCulloch and Dubovi EJ, (1996) Cross-sectional evaluation of environmental, host and management factors associated with risk of seropositivity to *Ehrlichia risticii* in horses of New York State. *American Journal of Veterinary Research* **57**: 278-285
- Bailey CJ, Reid SWJ, Hodgson DR, Suann CJ and Rose RJ, (1997) Risk factors associated with musculoskeletal injuries in Australian Thoroughbred racehorses. *Preventive Veterinary Medicine* **32**: 47-55
- Cohen ND, Peloso JG, Mundy GD, Fisher M, Holland RE, Little TV, Misheff MM, Watkins JP, Honnas CM and Moyer W, (1997) Racing-related factors and results of prerace physical inspection and their association with musculoskeletal injuries incurred in Thoroughbreds during races. *Journal of the American Veterinary Medical Association* **211**: 454-463
- Frankena K and Thrusfield MV, (1997) Basics of observational studies. In Application of Quantitative Methods in Veterinary Epidemiology Noordhuizen JPTM, Frankena K, van der Hoofd CM and Graat EAM Wageningen Pers, Wageningen
- Hillyer MH and Mair TS, (1997) Recurrent colic in the mature horse: A retrospective review of 58 cases. *Equine Veterinary Journal* **29**: 421-424.
- Martin SW, (1990a) Observational study methods and measures of association. In *Epidemiological skills in animal health*. Postgraduate Committee in Veterinary Science, University of Sydney pp57-64.
- Martin SW, (1990b) Sampling Populations Postgraduate Committee in Veterinary Science, University of Sydney pp 49-55.
- Maxson AD and Beef VB, (1997) Bacterial endocarditis in horses: ten cases (1984-1995). *Equine Veterinary Journal* **29**: 394-399.
- Reeves MJ, Salman MD and Smith G, (1996) Risk factors for equine acute abdominal disease (colic): Results from a multi-center case-control study. *Preventive Veterinary Medicine* **26**: 285-301
- Ross WA and Kaneene JB, (1995) A case-control study of an outbreak of Eastern Equine Encephalomyelitis in Michigan (USA) equine herds in 1991. *Preventive Veterinary Medicine* **24**: 157-170
- Slater MR and Hood DM, (1997) A cross-sectional epidemiological study of equine hoof wall problems and associated factors. *Equine Veterinary Journal* **29**: 67-69.
- Smith RD, (1995) *Veterinary Clinical Epidemiology A problem-oriented approach* 2nd Edition CRC Press
- Storgaard Jorgensen H, Proschowsky H, Falk-Ronne J, Willeberg P and Hesselholt M, (1997) The significance of routine radiographic findings with respect to subsequent racing performance and longevity in Standardbred trotters. *Equine Veterinary Journal* **29**: 55-59.
- Thrusfield MV (1995) *Veterinary Epidemiology*. 2nd Edition Blackwell Science
- Tinker MK, White NA, Lessard P, Thatcher CD, Pelzer KD, Davis B and Carmel DK, (1997) Prospective study of equine colic risk factors. *Equine Veterinary Journal* **29**: 454-458.

HOW TO DESIGN A SUCCESSFUL EQUINE EPIDEMIOLOGICAL RESEARCH STUDY

Roger Morris and Nigel Perkins

INTRODUCTION

A successful research study draws on a wide array of differing skills. These include creativity in formulating research questions, selection of an appropriate and cost-effective research design, definition of data gathering and data management methods, and making sure that the analytical approach chosen will provide maximum valid insights. Not only do we need to find solutions to all of these issues, but we also need to exercise good judgement in bridging the requirements of the ideal scientific design with the resources we are able to devote to the task.

In designing epidemiological studies, we all can sympathize with the frustrations outlined in Finagle's Laws:

- The information you have is not what you want
- The information you want is not what you need
- The information you need is not what you can get
- The information you can get costs more than you want to pay

Nevertheless, we can design and conduct valid and successful studies, provided that we plan the study effectively, have adequate numbers to ensure that it is interpretable, and carry it out in a way which provides sound results. This paper will examine each of the issues involved in producing sound results.

1. DEFINING THE RESEARCH QUESTION

The first step – and usually the one given least time and attention, is to identify the objective, and lay out what outcomes are to be considered in the study.

Figure 1.1 Design & implementation of a research project (Hulley & Cummings, 1988).

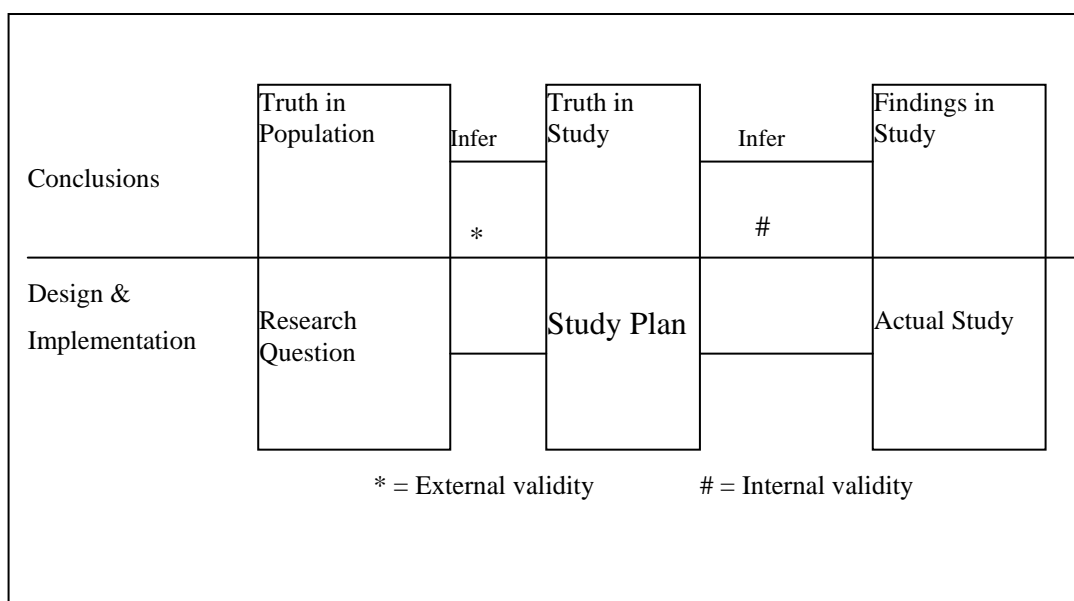


Figure 1.1 presents a diagrammatic representation of the research process. The research question is the objective of the study and is identified by considering the entire population. The researcher then designs a study which will allow valid conclusions to be drawn about the study sample, and then extended by epidemiological inference to the population of interest. The design and implementation phases move in the opposite direction to the inference phase.

One of the most common failings in research studies is to give inadequate time and thought to setting the objective clearly, in a way which allows it to be achieved with confidence that the conclusions are robust. This is particularly true for epidemiological research in which observational methods rather than designed experiments are the main approach adopted. It is critical to successful epidemiological research that the objective is defined clearly and written down. The outcome can then be measured against that definition, and judgments made about whether it has been achieved. This forces crucial rigour on the whole process, and avoids the risk of confusing structured field investigation with “stamp collecting”.

Figure 1.2 outlines the general series of decisions which should be considered in the design phase of a study. The remainder of this paper will be devoted to outlining the steps to be taken at each stage of designing the study, with special emphasis on issues relating to statistical significance, power of the study design and sample size estimation to achieve adequate power. These 3 issues are closely related and an understanding of statistical significance and power is critical to the estimation of sample size required for an experiment.

Figure 1.2 Study design (modified from Hulley & Cummings, 1988)

- | | |
|----|--|
| 1. | Research question |
| 2. | Epidemiological approach
Eg Observation vs Experiment |
| 3. | Subjects
Selection criteria
Sampling design |
| 4. | Variables
Independent (predictor variables)
Dependent (outcome variables)
Measurement methods |
| 5. | Data gathering and management |
| 6. | Statistical issues
Hypotheses
Sample size estimation
Analytical approach |

2. EPIDEMIOLOGICAL APPROACH

The core concept in the epidemiological approach is to study the population in its natural environment, and to view it as a system, in which diseases are produced by a causal web of interacting factors, rather than a linear process. These “risk factors” may be predisposing factors, enabling factors, precipitating factors or reinforcing factors, and some may have more than one of these functions.

In contrast, in the experimental approach most factors are held constant while one or two factors are varied over a range chosen by the experimenter. The two techniques are complementary, and both are needed. The experimental approach offers “strong inference” about the specific issue under evaluation, but may provide weak inference to the population because the experimental situation created may not have been close to the true causal web operating in the field. “Holding other factors constant” is simultaneously a blessing and a curse. Studying horses on a treadmill may give insights into mechanisms by which musculoskeletal stress can occur, but may tell us little about how it actually occurs on the racetrack, where interactions among factors are the key

to producing injury, and to preventing it.

The epidemiological method overcomes these limitations by researching the system in its true context, but because multiple factors may vary we are reliant on having included the important factors in our study. If we are not measuring the factors of importance then effects may be found to be significant but attributed to the wrong variable, or real effects may be masked. This problem is called *confounding*, and epidemiological methods focus very strongly on minimising this problem and maximising the inferential validity of the study.

There has been a very strong shift in recent years towards an epidemiological approach, as we have investigated more complex diseases, and as investigational methods and particularly analytical methods have been developed which allow confounding to be effectively controlled and multifactorial problems to be explored in their real-life context. Experienced researchers now use experimental methods to characterise pathogenetic mechanisms for disease and to assess the merits of control measures, in combination with epidemiological methods to resolve which risk factors are important, and how best to modify them.

3. SUBJECTS

In any epidemiological study, we want to draw sound *conclusions* about the *sample* of animals investigated, and sound, repeatable *inferences* about the *population of interest*, from which the sample was drawn. One of the important requirements is the selection of the subjects of the study, and the sampling design used to achieve that. In principle, we want the subjects to be a *random sample* of the population about which inferences will be made. Underlying statistical handling of data is this requirement that animals be a random sample, not a “representative” sample.

We usually design studies so that we subdivide our population of interest into sub-groups, before selecting randomly from the sub-groups. The *classification variables* we use to define the sub-groups (age, sex, case/control status etc) are crucial to making comparisons and are valuable tools for removing bias or confounding which would otherwise occur in the data.

The sampling design adopted will depend firstly on the study design we choose (which is covered in other papers) and secondly on other factors we might need to account for in a non-random way (eg by controlling for them). The sampling design will also need to take account of the need to have adequate power in the study, to be discussed below.

4. VARIABLES

Typically in an epidemiological study we evaluate the effect of a number of independent or predictor variables on one or more dependent (outcome) variables. Some predictor variables will be evaluated because we want to assess their effect, others to control their effect, and yet others as surrogate or proxy variables for factors which are of interest but not possible to measure (for example, management skill of the owner). In choosing variables to measure, it is important to identify ones which are as close to the true factor of interest as possible, and to avoid including two or more factors in an analysis which effectively measure the same item. This is as much an art as a science, and one which relies on experience and judgment to do well.

In measuring a variable, as a general rule the measurement should be made with maximum “information content”, which depends on how accurately the measurement can be made, plus how repeatable the measurement is. In principle, numerical measurement (interval or ordinal variable) is better than classification (categorical or binary variable), and if necessary the measurement can then be collapsed into a simpler scale.

5. DATA GATHERING AND MANAGEMENT

An important part of study planning is to develop the recording forms or systems which will be used. Traditionally this has required development of a range of recording sheets, and almost without exception researchers take up to 5 versions and three months before the forms are adapted to the task, and simplified down to what is realistic to gather. In future, palmtop recording devices such as the Palm Pilot will progressively take over many of the functions, and will assist in improving reliability of recording.

As a general rule, people collect far too much information, and lower the quality of what they get by diluting the effort. Never collect data without first working out how it will be used, and it is wise to mock up the analysis layout and plan before starting data collection – it soon becomes apparent that much of the data will never be used and can be cut out.

The important issues in data handling are to get the actual information from the animal to the computer file with minimum errors, and to ensure for longitudinal studies that there is sequential consistency between data over the series of observations – so that differences are true effects, not data integrity problems. Experience and research show that most of the errors in research data are in items for which the user has greater freedom to make an interpretation of evidence, whereas errors are low for items which are directly measured and recorded, with little contribution of personal judgment. Where possible, therefore, data should be directly recorded. Laboratory results also require much tighter data integrity checks than field measurements – which go straight into the file and leave little scope for later human error or test problems. There is a tendency to treat laboratory results as superior information, when in fact field data is inherently more reliable. Data quality checks should be designed in to study procedures, so that quality assurance actions become part of the study. If necessary, a particular piece of information may be gathered by two different techniques, and a decision made later about which to use. In many cases imperfect data on some points may be quite good enough for the particular study, while other data items may be worth considerable effort to measure accurately.

Where data must be gathered from questioning people rather than direct measurement, questionnaire design is critical to the validity of the findings. There are many tricks in developing successful questionnaires, and even more traps for the novice. Development of questionnaires requires considerable time and thought, plus extensive pre-testing.

6. STATISTICAL ISSUES IN STUDY MANAGEMENT

There is no point in spending effort on conducting a study if ultimately it is of questionable value. The design in combination with the number of observations made determine the statistical power of the study – which in simple terms can be seen as its capacity to detect an effect if one is there. In order to examine a study design, some principles of statistical inference need to be understood.

P values

When experimental results are analyzed there are two broad reasons for an apparent difference between groups of samples or subjects: the groups represent samples from identical populations and any observed difference in the actual samples measured is a coincidence (ie due to chance variation); or, the groups represent samples from populations which are in fact different. Statistical significance testing answers questions of the following type:

Figure 6.1 Calculation of the P value in testing H_0

If the populations from which the sample groups are drawn are in fact identical, what is the probability that the difference observed between the means of randomly selected subjects will be as large or larger than the actual difference observed in the study ?

The answer is a probability : the P value

This example uses a simple comparison of the means of 2 groups for the illustration of general principles. It is important to be aware of the assumptions underlying statistical testing. In this case it is assumed that samples are randomly selected and are representative of the populations from which they are drawn and that the experimental design is without flaws or biases.

The basis for statistical testing is commonly the hypothesis that the populations are the same and that any difference observed is purely due to chance. This is termed the null hypothesis (H_0). The alternative hypothesis states that there is a difference between the groups. Assuming the null hypothesis is true, statistical testing can be used to calculate the likelihood of observing various possible results. The proportion of these possible results where the difference between groups is as large or larger than the difference actually observed in the study, can then be calculated as a probability. This probability is termed the P value.

Figure 6.2 Example of a t test comparison of two means

Example 1	
$H_0 =$ Group A = Group B	$H_a =$ Group A \neq Group B
Group A	Group B
120	106
121	130
122	146
123	124
124	114
125	<u>113</u>
Mean 102	122.3
SD 11.1	14.3
Comparison of two means using an unpaired, 2-tailed t test produced a P value = 0.021	

In Figure 6.2, if the null hypothesis were true and the two samples were drawn from identical populations then we could expect to observe a difference as large as the one observed in this data set in about 2.1% of experiments. Alternatively about 97.9% of experiments would produce a difference smaller than the difference observed in this set of data (if the two populations were identical).

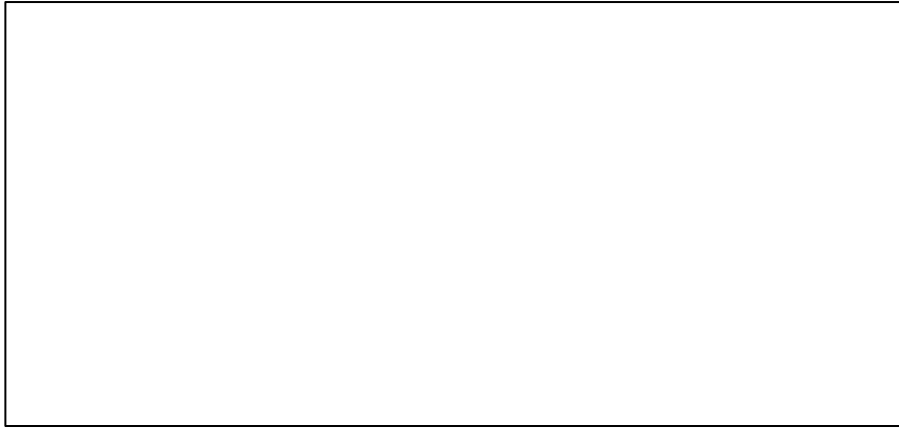
Statistical Significance

Statistical significance testing traditionally involves the following steps:

1. State the null hypothesis
2. Define a threshold value for declaring a P value significant. This threshold value is commonly 0.05 and is termed the significance level or α .
3. Select an appropriate statistical test and calculate the P value.

4. If the calculated P value is less than the defined threshold the difference is defined as significant and the null hypothesis is rejected. If the calculated P value is greater than α , the difference is said to be not significant and the null hypothesis is accepted.
5. If the null hypothesis is rejected then the alternative hypothesis is accepted.

Figure 6.3 Errors in statistical significance testing



The use of 0.05 as the threshold of significance is traditional but arbitrary. If $\alpha = 0.05$ then the analysis will have a 5% chance of incorrectly rejecting the null hypothesis. Commonly used levels include 0.001, 0.01, 0.05. Different definitions of significance level should be considered depending on the particular research project being planned. As α is lowered, type I errors (mistakenly rejecting H_0 when the groups are actually not different) become less likely but type II errors (missing a real difference), become more likely. An example of an experiment using a relatively high value of α might be preliminary testing of a new drug. Under these circumstances it may be preferable to set a high α and low β in order to minimize the chance of discarding a drug which might have an effect.

There have been two important trends in recent years in the reporting of statistical significance. These are the reporting of actual P values rather than <0.05 or >0.05 and the use of confidence intervals in significance testing. The practice of reporting actual P values should be encouraged particularly since software programs currently used for statistical significance testing can easily calculate the exact P value. For example, using $\alpha=0.05$, a calculated P value of 0.049 could be reported as $P<0.05$ and a calculated P value of 0.051 as $P>0.05$. One is significant “by definition” and the other not - and yet the difference between the two is likely to be trivial. Reporting the actual P value allows the reader to appreciate such details and may increase the usefulness of the research results.

Confidence intervals are gaining strong support as a means of presenting study estimates of population values. Confidence intervals give an indication of the variability the estimate would have if the experiment were to be repeated a large number of times. Confidence intervals define an upper limit and a lower limit with an associated probability. For example measuring variable x in a sample of 22 subjects produced Mean = 43.00, SEM = 4.90 and SD = 23.00. The 95% confidence interval for this mean is 32.80 to 53.20. This is interpreted as indicating that there is a 95% chance that the range 32.8 to 53.2 will include the true population mean for variable x. Calculations of P values and confidence intervals (CIs) are based on the same statistical principles and assumptions, and are closely related.

If the 95% CI includes the value stated in the null hypothesis then the P value must be greater than 0.05 and the null hypothesis is accepted. For example in a comparison of two means, if the 95% CI of the difference between the two means includes 0, then the null hypothesis is accepted. If the 95% confidence interval for the relative risk in a prospective study includes 1, then the P value for the relative risk being different from 1.0 must be greater than 0.05.

Confidence intervals can also be used to estimate required sample size for a given experiment. Under most circumstances, increasing sample size will make the CI narrower. If a desired CI can be identified then the number of subjects required to produce the desired width CI can be estimated.

Confidence intervals can therefore be used to provide the same information that a traditional statistical test does when it produces a P value. Confidence intervals provide additional information that the P value cannot provide in defining the variability of estimates based on the experiment. Confidence intervals are gaining favour in hypothesis testing because they clearly illustrate the magnitude of the observed difference. A 95% or 99% confidence interval will more comprehensively describe the degree of confidence which can be attached to a particular finding than a simple $P < 0.05$.

Power

The probability of making a type II error is termed beta (β) and $[1 - \beta]$ is termed *Power*.

Figure 6.4 Power (Thomas & Krebs, 1997)

Power = the probability of obtaining a statistically significant result in a study given that there is a biologically real effect in the population being studied.

$$\text{Power} = [1 - \beta]$$

Beta is commonly set within the range of 0.05 to 0.2. If $\beta = 0.1$, the analysis has a 10% chance of missing a real association or difference. Alternatively this can be described as a power of 90% which implies that the analysis has a 90% chance of detecting an effect if an effect exists in the population. Describing a study as having a power of 90% means that if the study were to be repeated a large number of times, 90% of the repetitions would be expected to report a significant P value if there is a true difference. Increasing power is desirable since it will improve the chance of detecting an effect, but it must be balanced against increased cost and some risk of claiming an effect when there is no true difference. The logical ways of increasing statistical power in an experiment are by:

- I. increasing sample size (n)
- II. increasing effect size or choosing a test/variable which is associated with a larger effect size.
- III. Increasing α .

Effect Size

The likelihood of detecting an association between an independent and a dependent variable or of finding a significant difference in a statistical test is also dependent on the magnitude of the association or difference. The magnitude of the difference or association in a statistical sense is termed the *effect size*. If furosemide treatment genuinely reduced the risk of exercise-induced pulmonary haemorrhage by 90%, a relatively small sample size in an experiment would be able to detect the effect. If treatment only reduced bleeding by 2% a small sample may not detect the effect at all. The effect size in a study comparing two means is the difference between the means divided by the average standard deviations from the two groups. In a study comparing two proportions, the effect size is the difference between the two proportions divided by the pooled standard deviations (Dawson-Saunders and Trapp, 1994). By dividing the difference by the standard deviation, the resultant effect size is a standardized, unitless figure which makes it suitable for statistical calculations. When

performing hypothesis testing the null hypothesis is that the groups do not differ ie the effect size is equal to zero. When the null hypothesis is rejected and the groups are said to be different, then the effect size is a specific nonzero estimate which can be used as an index of the degree of departure from the null hypothesis (Cohen, 1988).

Effect size has an important influence on power analysis and sample size determination. The larger the effect size, the smaller the sample required to detect it with a given probability.

Power Analysis

This involves the combined use of statistical significance, power and effect size in planning experiments. The variables above form a closed relationship. Estimates of effect size, sample size or power can be calculated by inserting values for the other components of the equation. Depending on the type of statistical analysis necessary for the experimental data, an estimate of variation of data may also be necessary in power analysis. The planning phase of any experiment should include power analysis as outlined below.

Power analyses are based on estimates of underlying characteristics of the variables to be measured, and should not be considered precise measurements – especially since they are done before the study commences and the data is collected. Although α and β are usually set to conventional values (0.05, 0.2), researchers should consider defining α and β according to the specific needs of the experiment being planned. Effect size is typically defined as the smallest difference which would be scientifically or clinically meaningful. Estimating expected effect size and data

variability can be difficult. Potential sources of such information may include published literature, personal communications from researchers active or experienced in the field of study, pilot tests and ultimately a guess based on judgement. In some cases where estimating effect size or variability is very difficult, the variable can be dichotomized into “above the median” and “below the median” and power analysis performed on proportions (which does not require an estimate of effect size or variability).

Experimental studies

1. Define the level of significance (α) in relation to the null hypothesis
2. Estimate how high the chance should be of detecting an actual difference in the study (Power or $[1-\beta]$).
3. Select appropriate statistical tests based on variables and study design
4. Estimate the smallest size of the difference between means which could be considered to be of clinical importance (effect size). Alternatively depending on the nature of the study, estimate the smallest magnitude of association between variables which could be considered meaningful.
5. Estimate the standard deviation in the population (estimate of variation in data).
6. Estimate required sample size using a formula dependent on the type of statistical test being used to analyse data.

Power analysis serves the following functions:

1. Estimate sample size required to detect an estimated effect size having set α and β . Performed during the planning phase of an experiment.
2. Post hoc power analyses (performed after completion of the study):
 - a. estimate sample size which would have been required to detect an effect size of the magnitude observed in the study, given a defined α and β .
 - b. estimate the smallest effect size which could be detected given a defined sample size, α and β .
 - c. estimate the power of the experiment as it was performed, which helps show whether a significant result was even possible, given the real values

found in the study for the various factors of importance.

Minimizing Sample Size

Using power analysis to calculate required sample size may result in an estimate which exceeds practically possible limits. Under these circumstances it is necessary to perform a series of analyses using different scenarios. The following changes may be used in power analysis to lower required sample size:

1. Reduce defined effect size.
2. Increase α and β .
3. Change the analytical approach to use a statistical procedure which has inherently greater power.
4. Reduce the necessary confidence level.
5. Use continuous variables instead of dichotomous variables. This permits a smaller sample size for a given power or a greater power for a given sample size.
6. Increase precision of data. It may be possible to measure a different variable with more precision, or use a different test to measure the same variable(s). Reducing data variation will also allow a smaller sample size for a given power.
7. Use paired measurements. In certain study designs it may be possible to alter the design slightly to involve paired measurements (before and after). This reduces data variation.
8. Use unequal group sizes. In most circumstances using equal group sizes gives the greatest power for the total number of subjects. In some cases it may be easier or cheaper to increase the size of one group and not the other. Many case/control studies involve control groups which are considerably larger than the case groups.
9. Measure variables which are more common in the study sample.

The strategies listed above are examples which may be used to maximize the chance of a study producing results which include statistically significant and meaningful results. Some involve making changes to study design and these must be considered with caution since they may actually alter the focus of the research objectives.

Where To Find Formulae, Tables And Software

Simple estimates of required sample sizes for some tests can be obtained using published tables. These can be found in a variety of texts including Cohen (1988), Hulley and Cummings (1988), Thrusfield (1995). A reasonably complete offering of formulae used for power analyses can be found in Thrusfield (1995).

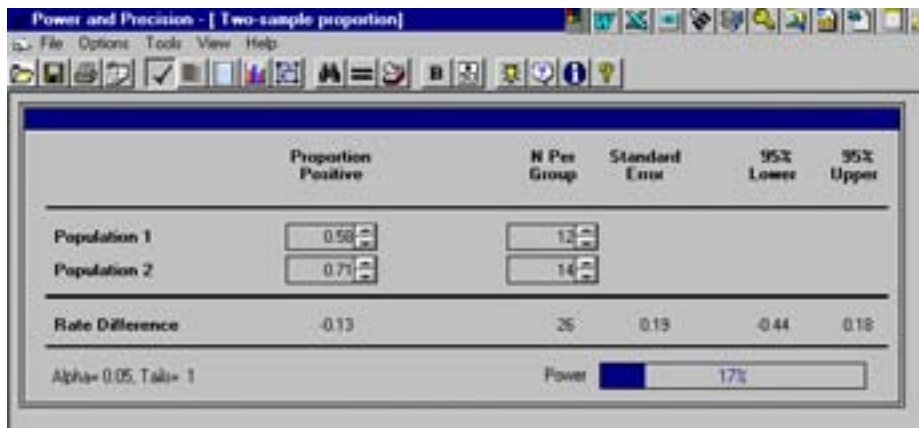
A large number of computer software programs offer varying capabilities of power analyses. Many general purpose statistical programs include some power analysis capabilities. There are also a reasonably large number of specialized software packages which concentrate solely on power analysis routines. A comprehensive listing of power analysis software and a recent review of available software may be found on the world wide web at the following http address:

<http://www.interchg.ubc.ca/cacb/power/review/powrev.html>

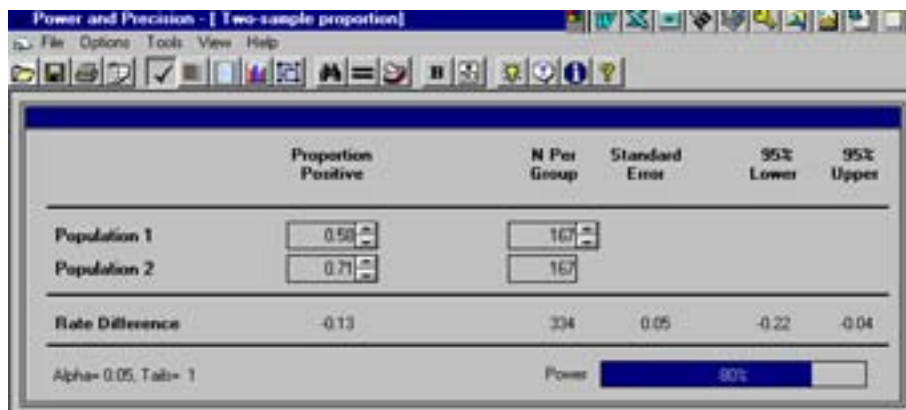
<http://www.interchg.ubc.ca/cacb/power/>

Example: Pregnancy rates in mares bred at the first or second heat post foaling

This pilot study involved using 100 mares at a research stud farm to investigate the premise that selecting mares to be bred at foal heat was as successful as breeding mares at the second heat post foaling. Mares were randomly assigned to groups after foaling and were either bred at the first heat post foaling (foal heat) or were allowed to ovulate and then received PGF2alpha and were bred at the second ovulation. Between the time the study was designed and its completion, a major source of funding for the farm ended and mare numbers dropped to 26 (from over 100). Although the low numbers meant a meaningful result was unlikely, the study was completed with the observed pregnancy rates being 58% and 71% in the foal heat and second heat groups respectively. Group sizes were 12 and 14 mares. Comparison of the data using a Chi square analysis revealed no significant difference between the two management methods ($P=0.595$ using a Fisher exact 1-tailed test). Retrospective power analyses were then performed to provide some additional information. The window below shows the data entered into the two sample proportion window of Power and Precision with the calculated power being equal to 0.17. This means that using this number of animals and the difference in pregnancy rates, a significant difference could be expected in 17% of experiments.



The same program can then be used to estimate the number of animals which would be required to return a significant result given the same pregnancy rate as that observed above. For this calculation the power is defined as 0.8 and $\alpha = 0.05$.



From this table it can be seen that in order for a 13% difference in pregnancy rate to be detected as a significant difference with power = 0.8 and $\alpha = 0.05$, the experiment required 167 animals in each group.

Using a similar program (Pass 6.0) we can then perform additional scenarios. Assuming that the pregnancy rate observed at the second ovulation breeding (71%) was representative of the expected population rate, what would be the minimum difference which would be deemed significant given the following assumptions (power =0.8, $\alpha=0.05$, with varying numbers of mares in the experiment). This is detecting the minimal effect size and Pass returns the following output:

Two Proportions Power Analysis							
Page	1						
Time/Date	12:56:49 03-20-1998						
Numeric Results with Ha: P1<P2							
Power	N1	N2	P1	P2	Alpha	Beta	
0.00000	25	25	1.000000	0.710000	0.050000	1.00000	
0.80000	50	50	0.468100	0.710000	0.050000	0.20000	
0.80000	100	100	0.540860	0.710000	0.050000	0.20000	
0.80000	150	150	0.572910	0.710000	0.050000	0.20000	
0.80000	167	167	0.580320	0.710000	0.050000	0.20000	
0.80000	200	200	0.591870	0.710000	0.050000	0.20000	

From these results we can see that with 50 mares in each group and a second ovulation pregnancy rate of 71%, the pregnancy rate in mares bred at foal heat would have to be as low or lower than 46.8% in order for the result to be deemed significant.

CONCLUSION

The art of designing an epidemiological study which has a high probability of achieving its clearly defined objectives has advanced considerably in recent years, and may even be considered to be approaching a science. We have sufficient accumulated experience to avoid major pitfalls, and we now have tools such as power analysis software to decide the adequacy of a proposed design, and refine it to produce maximum information content for the research investment.

REFERENCES

Borenstein M, Rothstein H and Cohen J. Power and Precision. *Software program for Windows PC*. <http://www.PowerAndPrecision.com>

Cohen J., (1988) *Statistical power analysis for the behavioral sciences*. 2nd edition. Hillsdale, NJ. Lawrence Erlbaum Associates..

Dawson-Saunders B and Trapp RG., (1994) *Basic and Clinical Biostatistics*, 2nd edition.. Prentice Hall International Inc, NJ..

Gardner MJ and Altman DG., (1989) *Statistics with confidence – Confidence intervals and statistical guidelines*. British Medical Journal, London..

Helberg C. Pitfalls of data analysis (or how to avoid lies and dammed lies). 3rd Int Conf on Statistics in Industry, Dallas, TX, June 1995. <http://www.rdsu.wisc.edu/pitfalls/>

Hulley SB and Cummings SR., (1988) *Designing Clinical Research : An Epidemiologic Approach*. Williams & Wilkins, Baltimore USA,.

Pass 6.0 Statistical Software for Windows. NCSS Statistical Software, Kaysville Utah, USA. <http://www.ncss.com>

Thomas L and Krebs CJ. A review of statistical power analysis software. *Bull Ecol Soc America*, 78(2): In Press, 1997. And at <http://www.interchg.ubc.ca/cacb/power/review/powrev.html>

Thrusfield,M., (1995) *Veterinary epidemiology*. 2nd edition. Blackwell Science, Oxford, England, 479pp..

EQUINE MORBILLIVIRUS: AN EPIDEMIOLOGICAL PERSPECTIVE

Chris Baldock

SUMMARY

Two separate outbreaks of disease due to infection with equine morbillivirus (EMV) have occurred in Queensland. Although first reported as occurring in Brisbane, the first outbreak actually occurred near Mackay in northern coastal Queensland in August 1994 and involved the death of two horses and a male horse breeder. The second outbreak occurred in Brisbane in September 1994 where 14 horses and their trainer died and a further 7 horses and a stablehand were infected but survived. The 7 infected horses were later destroyed. Despite considerable effort, no link has yet been established between the two outbreaks.

Extensive follow-up investigations of in-contact horses as well as clinical and serological surveillance in horses and other animals following both outbreaks has revealed no further evidence of infection in Queensland. It is now thought that fruit bats are the natural hosts of the virus which caused the outbreaks of disease in horses and people. The virus isolated from fruit bats has been called bat paramyxovirus (BPV) and it appears to be identical to the virus originally called EMV. This virus has probably been in Australian bat populations for some time.

The evidence to date suggests that EMV is not endemic in the Queensland horse population, and that spillover from fruit bats is rare. In addition, in both outbreaks, transmission appears to have occurred only from the index case, suggesting that EMV is not easily transmitted among horses.

INTRODUCTION

The first descriptions of disease due to infection with equine morbillivirus (EMV) in man (Selvey and Sheridan 1994; Selvey *et al* 1995) and in horses (Murray *et al* 1995) arose from an outbreak which occurred in a Brisbane racehorse stable complex in September 1994. Subsequently, investigation of an acute encephalitis in a farmer in October 1995, revealed that a smaller outbreak of EMV had occurred near Mackay in north Queensland during August 1994 (Allworth *et al* 1995). These two incidents are the only known outbreaks of EMV infection to have occurred anywhere in the world.

Rarely is one confronted by a "new" disease which generates so much public interest. Equine morbillivirus, investigated since first being recognised in late 1994, offers many examples of epidemiology at work. For a number of reasons, there has been limited opportunity to study this disease in pen experiments, so much of what we know has come from field observations. These include:

- unravelling the natural history of a disease;
- deducing mechanisms of transmission;
- applying outbreak investigation techniques;
- making disease control decisions with imperfect knowledge;
- the use of unproven serological tests;
- developing sound surveillance strategies;
- developing investigation priorities;
- studying wildlife reservoirs.

In the case of the initial investigations of EMV in Brisbane, epidemiologists played important roles in not only describing the outbreak, but in quickly developing hypotheses on the most likely cause and mechanism of transmission so that disease control strategies had a rational basis despite the initial lack of an aetiological diagnosis. With EMV, there was also an urgent need to have a well designed approach to prioritisation of tracings and surveillance to ensure efficient use of resources. The Queensland Department of Primary Industries collaborated with veterinarians and scientists from a number of other agencies and individuals in the investigations and control of this new disease.

Although the text of this paper describes epidemiological aspects of EMV, the presentation will be in the form of an interactive class exercise, with participants playing the role of "population detectives". In an example such as EMV, the epidemiologist must take all the available information and develop the most consistent picture on which to base disease control decisions. Often, this information is incomplete and during the exercise, participants will learn how they can increase the likelihood of taking the correct decision in the face of uncertainty. The outbreaks are described in the chronological order of events rather than in the order of the investigations.

THE MACKAY OUTBREAK

Although this episode was first investigated in October 1995, the outbreak occurred in August 1994 on a property near Mackay in north coastal Queensland. The property, owned and managed by a farmer and his wife, a veterinary surgeon, was used for growing sugar cane and breeding thoroughbred horses.

On 1 August 1994, a heavily pregnant, 10 year old brood mare (horse A) died suddenly in a paddock. Reported clinical signs were severe respiratory distress, ataxia and swelling of the head particularly the area of the infraorbital fossa and the cheeks. A second horse, a two year old stallion (horse B) from an adjoining paddock was reported to have licked the carcass. This horse died on 12 August within 24 hours of becoming ill with signs of aimless pacing, muscle trembling and a haemorrhagic nasal discharge. Both horses were autopsied and specimens were sent to a private veterinary laboratory. Histopathological findings were inconclusive.

The farmer assisted with the treatment and subsequent autopsies of both horses. He became ill in August-September 1994 with a mild meningo-encephalitis which improved with antibiotic therapy. In mid-September 1995 the owner developed a severe encephalitis-like syndrome and died on 21 October 1995. He had IgG antibody to EMV and his CSF was positive to EMV on PCR (Allworth *et al* 1995).

The farmer's death led to a retrospective investigation of the horse deaths. Limited preserved material from both horses was still available in October 1995 and was tested for presence of EMV using PCR and immunostaining. The conclusion was that, in horse B, EMV was the cause of death and that horse A was infected but the cause of death could not be confirmed because of insufficient material for a histopathological diagnosis. From all the available evidence it is reasonable to assume that horse A was a case of EMV and that horse B was infected from horse A.

Horse A had been in the same paddock in the company of a number of other horses for two months prior to her death with nothing untoward happening. Approximately 10 days prior to her death she received tetanus toxoid, had her feet trimmed and was wormed.

Extensive serological surveillance was undertaken and the initial sampling completed by 23 November 1995 with negative findings. This surveillance included bleeding all other domestic animals on the affected and surrounding properties as well as horses which had spent any time on the property from July 1994. In addition, a serological survey of the Queensland paddocked horse population was undertaken with negative findings in 2024 animals tested (Ward *et al* 1996). Sampling of wildlife and some avian species was also

undertaken in an attempt to identify a possible reservoir host, again with negative findings (Rogers *et al* 1996). The overall serological surveillance effort for both the Mackay and Brisbane outbreaks is summarised in Table 1.

Table 1. Summary of total serological surveillance for equine morbillivirus with negative findings by serum neutralisation test from September 1994 to April 1995

Animal species	Number sampled
Horse - Mackay ¹	2869
Horse - Brisbane ²	1964
Donkey	1
Cattle, goat, pig	287
Chicken, duck, goose, turkey	33
Other birds	2
Cat	564
Dog	23
Cane toad, bandicoot, flying fox, kangaroo, mouse, snake, wallaby, other wildlife	166
Total	5909

1. Horses sampled in 1995 following detection of the Mackay outbreak.
2. Horses sampled in 1994 following the Brisbane outbreak.

THE BRISBANE OUTBREAK

In September 1994 a focal outbreak of an acute respiratory syndrome resulted in the death of 14 horses and their trainer (Selvey and Sheridan 1994). Presenting signs were suggestive of African Horse Sickness, a disease exotic to Australia, or a toxicosis. Subsequent investigations revealed the cause to be a hitherto undescribed morbillivirus now known as equine morbillivirus (Murray *et al* 1995).

The outbreak occurred in a stable complex in the suburb of Hendra in Brisbane. The putative index case was a pregnant mare which was ill on 7 September when moved from a spelling paddock in the suburb of Cannon Hill to the Hendra stables where 23 other horses were located.

A further 16 horses became ill between the 16 and 23 September and 12 of these died from several hours to 3 days after the onset of illness. Another horse was reported to have died on approximately 2 September in the Cannon Hill paddock but it could not be established whether or not it was a case of the same syndrome. Three other horses seroconverted but did not show demonstrable clinical signs. Two of the recovered horses developed mild neurological signs consisting of myoclonic twitches similar to those seen in some dogs after following infection with canine distemper virus. The signs were transient in one case but persisted until euthanasia in the other (Rogers *et al* 1996). The trainer and one of the other people in close contact with the putative index case also became ill and, despite intensive medical care, the trainer died. The incubation period in the natural cases in horses was mostly 8 to 11 days with a maximum possible of 16 days.

The pattern of the outbreak suggested that all other cases were a result of transmission from the putative index case. Natural transmission was most likely direct via frothy nasal

discharges as a consequence of close contact or mechanical transfer. Aerosol transmission seemed unlikely as sneezing and coughing were not a feature of the syndrome. Transmission did not appear to occur during the incubatory phase of the disease. All of the available evidence suggested that under the conditions of the outbreak, EMV was not highly contagious (Baldock *et al* 1995).

Quarantine and strict movement restrictions were used to contain the outbreak. Surveillance showed that the outbreak was successfully contained and was an isolated focal event. The surveillance strategy had three components: targeted clinical and serological surveillance through tracing of risky contacts; passive clinical surveillance based on reports from private practitioners and the general public; and a planned serological survey to ascertain whether or not infection with EMV was more widespread (Baldock *et al* 1995).

SIMILARITIES BETWEEN THE MACKAY AND BRISBANE OUTBREAKS

Despite considerable effort, no link has yet been established between the Mackay and Brisbane incidents. However, the fact that they occurred so close together in time in addition to other similarities suggests there may be some, as yet undiscovered, link. The points of similarity between the two incidents are worth summarising:

- both outbreaks occurred in a two month period in late winter to early spring, 1994;
- the index case was an older, heavily pregnant mare at pasture;
- the mare had been in the paddock for a duration well in excess of the incubation period observed in other horses;
- other horses which were in the paddock with the index case remained well and did not seroconvert;
- transmission occurred from the index case but there was apparently no secondary transmission in horses;
- infection appears to have been transmitted from a horse to a person.

CAUSE OF THE OUTBREAKS IN HORSES

This was a key issue from the outset of the Brisbane outbreak investigations beginning on 22 September 1994. Prior to the establishment of a viral aetiology, control strategies were based on the assumption that the most likely cause was a directly transmitted agent with a toxic aetiology almost as likely and a vector borne disease quite unlikely.

The pyrexia in clinical cases initially led to support for an infectious rather than toxic aetiology. Also, the very tight spatial and temporal clustering of cases led to the view that direct transmission was more likely than involvement of a vector. Initial histopathology showed damage to endothelial linings to small pulmonary blood vessels leading to pulmonary oedema. This damage was hypothesised to provide a route of exit for an infectious agent and it was therefore assumed that the frothy nasal discharges were infectious.

The initial hypothesis that the outbreak at the Hendra stables originated from the mare moved from Cannon Hill on 7 September was based on the similarity in clinical signs and the fact that all other cases at Hendra developed signs after this case including the two human cases. Because these dates were known precisely, estimates of the incubation period could be made. A horse which was moved from the Hendra stables on 9 September subsequently seroconverted to EMV. This horse left the stables before any other horses or the humans became ill. This evidence gave further credence to the role played by the mare introduced from the Cannon Hill paddock.

Laboratory investigations subsequently revealed the aetiology to be EMV, a previously unrecorded virus (Murray *et al* 1995).

NATURAL HISTORY AND TRANSMISSION IN HORSES

The incubation period in the natural cases was mostly 8 to 11 days with a maximum possible of 16 days. In transmission tests carried out at the Australian Animal Health Laboratory (AAHL), the incubation period was 3 to 10 days.

All the other cases (20) in horses and the 2 suspect human cases can be linked to the original index case which was first observed sick on 7 September and died on 9 September.

In horses, EMV is not highly infectious and requires direct contact for natural transmission to occur, possibly involving nasal discharges which occur in the terminal stages of the disease. The microscopic pathology of affected lungs indicates that such discharges are likely to contain large quantities of virus. In the Mackay incident, the route of infection for horse B was most likely either oral or nasal since it was reported as having licked the dead mare through a fence. The evidence for a specific route of infection in the Brisbane outbreak is not as strong but a naso-oral route is probable. In both outbreaks, evidence indicated that transmission only occurred from the index case with no secondary transmission. In addition, fully susceptible horses which were in contact with cases did not become infected. Aerosol transmission seemed unlikely as sneezing and coughing were not a feature of the syndrome. The possibility of iatrogenic transmission was investigated but there was no supportive evidence. However, iatrogenic transmission remains consistent with the pattern of infection in the Brisbane outbreak.

Transmission did not appear to occur during the incubatory phase of the disease. At Hendra, horses which subsequently succumbed to the disease were freely mixing with the wider horse population while incubating and did not transmit the virus to the best of our knowledge.

All of the available evidence suggested that under the conditions of the two known outbreaks, EMV was not highly contagious. The disease did not spread from secondary cases despite ample opportunity.

HOST RANGE

The two outbreaks have shown that both horses and humans are susceptible. Because of the risk to humans, EMV is regarded as a biohazard level 4 agent requiring maximum biosecurity. The only veterinary laboratory in Australia with a capacity to work with such agents is the Australian Animal Health Laboratory at Geelong (AAHL).

Limited experimental studies at AAHL have shown that cats and some guinea pigs succumb to the disease following inoculation with large doses of the virus. Affected animals showed respiratory signs in the terminal stages with pulmonary lesions similar to those occurring in horses (Westbury *et al* 1995).

Mice, rats, rabbits, chickens and dogs did not show clinical signs, there were no gross or microscopic lesions evident 21 days after challenge, virus could not be isolated and mice and chickens did not seroconvert. However, equivocal neutralising antibody titres were detected in some rats and one of two inoculated dogs while the two inoculated rabbits had significant titres of 1:2560 and 1:320 (Westbury *et al* 1995).

More recent work has shown that the natural hosts for this virus are fruit eating bats - their story is told below.

FRUIT BATS - A NATURAL WILDLIFE RESERVOIR

Following each of the two outbreaks, wildlife trapping programs were undertaken in the vicinity of each index case in the search for a possible wildlife reservoir (Young *et al* 1996). Initial trapping was not targeted to any particular type of animal but included only ground dwelling animals as well as insects. No evidence of EMV infection was found in over 30 species of rodents, marsupials, birds, amphibians and insects tested. However, very few

animals of any particular species were caught and all were negative on serology (Rogers *et al* 1996). The small sample size for any particular species however, limited meaningful interpretation and indicated the need for a more targeted approach to wildlife sero-surveillance.

A timely comparison of PCR products from both outbreak locations indicated that virus from both sites was identical. In the absence of any apparent epidemiological connection between the two sites, the finding suggested a source of infection that was common to both locations. Consequently, a "think tank" comprising experts in a wide range of disciplines was convened in January 1996 for the purpose of establishing priorities for wildlife sampling. A set of criteria which a potential reservoir species would need to meet were developed based on these observations that the Mackay and Brisbane outbreaks appeared to be separate events yet the viruses isolated from each were genetically identical. These criteria were:

- the species should be present in both Mackay and Brisbane;
- the species was capable of migrating between the two outbreak sites or at least have overlapping and mixing populations which spanned the two locations; and
- contact with horses should be possible.

Two broad groups of animals met these criteria - migratory birds and Pteropus bats (commonly known as flying foxes or fruit bats). However, because EMV was a mammalian virus and transmission of paramyxoviruses from birds to mammals is rare, bats were given a higher priority (Young *et al* 1996). In addition, a paramyxovirus (Mapuera virus) had recently been described in fruit eating bats in Brazil and India. The wildlife program was subsequently redirected towards Pteropus bats.

In Australia there are four species of Pteropus bats: the spectacled (*Pteropus conspicillatus*), the black (*Pteropus alecto*), the little red (*Pteropus scapulatus*) and the grey-headed fruit bat (*Pteropus poliocephalus*). All four species are present in coastal Queensland.

Opportunistic sampling of sick or injured wild flying foxes in temporary captivity with veterinary practitioners and wildlife carers was employed as a means of preliminarily sampling wild populations. While such samples were recognised as potentially biased, they were seen as adequate for screening purposes.

In April 1996, antibodies to EMV were found in all four species of flying foxes occurring in Australia. These are the little red (*Pteropus scapulatus*), the spectacled (*P. conspicillatus*), the black (*P. alecto*), and the grey headed flying fox (*P. poliocephalus*). The overall seroprevalence was estimated between 10 and 15% (Field *et al* 1997).

In September 1996, two years after the first reported outbreak of EMV, a paramyxovirus, tentatively named bat paramyxovirus (BPV), was isolated from the uterine fluid of an apparently healthy grey-headed flying fox which aborted twin foetuses after injuring herself on a wire fence.

This isolate returned positive results by immunofluorescence using both human and horse serum positive for EMV antibodies. Both BPV and EMV produced similar changes in cell culture. Both appeared identical by electron microscopy. A constant serum-varying virus neutralisation test showed that EMV antibody positive horse serum completely neutralised the BPV isolate. Sequencing of a 200 base pair fragment of the matrix protein gene has shown that the BPV sequence is identical to the EMV sequence. These results all indicate that BPV and EMV are one and the same virus. Isolates of BPV have now been obtained from three species of flying foxes. BPV has been isolated from blood, kidney, uterine fluids, and foetal tissues.

The wide geographical distribution in a number of flying-fox species suggests that BPV has been in Australian flying fox populations for some time. In addition, several Australian bat species have geographic distributions which extend beyond Australia. For example, the

range of the black flying fox *P. alecto* extends to New Guinea and the Indo-Malay peninsula. Recent serological studies suggest that BPV also occurs in flying foxes in Papua New Guinea - there has been positive serology in 4 additional *Pteropus* species (*P. neohypernicus*, *P. capistratus*, *P. hypomelanus*, *P. admiralitatum*) and 2 *Dobsonia* species (*D. mollucense*, *D. andersonii*). *D. mollucense* also occurs in far north Queensland (Field *et al* 1997).

Despite extensive testing, there has been no evidence of BPV in insectivorous (micro) bats in Australia.

DISCUSSION

The characteristic and severe nature of EMV infection as well as the lack of evidence for exposure in the Queensland horse population initially suggested that horses were not a maintenance host for EMV. Although a wildlife reservoir was hypothesised from the outset, initial investigations proved fruitless. It was not until the Brisbane and Mackay incidents were reviewed that a more targeted approach was taken and was quickly successful.

It is now reasonable to assume that BPV and EMV are the same virus and that fruit bats are natural hosts for BPV which has been in Australia for some time. It would also appear that spillover to horses is rare and outbreaks are short and self limiting. However, the circumstances which produced two incidents in two consecutive months in 1994 remain a mystery. At this time, it is not known how horses are infected from bats, but it is likely that the three known human cases have been infected from horses and not from bats directly. Also, the significance of the infection in bats remains unknown as do mechanisms of transmission from bat to bat.

REFERENCES

- Allworth T, O'Sullivan J, Selvey L, Sheridan J., (1995). Equine morbillivirus in Queensland. *Comm Dis Intell* **19**: 575.
- Baldock FC, Pearse BHG, Roberts J, Black P, Pitt D, Auer D., (1995). Acute Equine Respiratory Syndrome (AERS): The role of epidemiologists in the 1994 Brisbane outbreak. *Epidemiology Chapter Proceedings, Australian Veterinary Association Annual Conference*. Ed. J. Morton. Australian College of Veterinary Scientists. pp 174-177.
- Field H, Halpin K, Young P., (1997). Emerging viruses of Australian bats - early epidemiological indications. *Proceedings of Epidemiology Program, 10th FAVA Congress*, 24-28 Aug 1997, Cairns, Qld. pp 255-259.
- Murray PK, Selleck P, Hooper P, Hyatt A, Gould A, Gleeson L, Westbury H, Hiley L, Selvey L, Rodwell B, Ketterer P., (1995). A morbillivirus that caused fatal disease in horses and humans. *Science* **268**: 94-97
- Rogers RJ, Douglas IC, Baldock FC, Glanville RJ, Seppanen KT, Gleeson LJ, Selleck PN, Dunn, KJ., (1996). Investigation of a second focus of equine morbillivirus infection in coastal Queensland. *Aust Vet J* **74**: 243-244.
- Selvey L, Sheridan J., (1994). Outbreak of a severe respiratory disease in humans and horses due to a previously unrecognised paramyxovirus. *Comm Dis Intell* **18**: 499.
- Selvey L, Wells RM, McCormack JG, Ansford AJ, Murray PK, Rogers RJ, Lavercombe PS, Selleck P, Sheridan JW., (1995). Infection of humans and horses by a newly described morbillivirus. *Med J Aust* **162**: 642-645.
- Ward MP, Black PF, Childs AJ, Baldock FC, Webster WR, Rodwell BJ., (1996). Negative findings from serological studies of equine morbillivirus in the Queensland horse population. *Aust Vet J* **74**: 241-243.
- Young PL, Halpin K, Selleck PW, Field H, Gravel JL *et al.*, (1996). Serological evidence for the presence in Pteropus bats of a paramyxovirus related to equine morbillivirus. *Emerging Infect Dis* **2**:239-240.
- Westbury, HA, Hooper, PT, Selleck, PW and Murray, PK., (1995). Equine morbillivirus pneumonia: susceptibility of laboratory animals to the virus. *Aust Vet J* **72**: 278-279.

WASTAGE IN THE RACING INDUSTRY - APPROACHES TO STUDY

Craig J. Bailey

INTRODUCTION

The term “wastage” has been used to refer to losses that occur at all stages of development of a racehorse, including mares failing to conceive or not carrying the pregnancy to term, perinatal mortalities, horses not entering training or not racing, injury-induced delays in training, and the premature end of racing careers due to serious injury or fatality. In the United Kingdom, it was calculated that 72.8% of applicable covered returned Thoroughbred mares did not produce a foal, or produced a foal that did not race by four-years of age (Jeffcott *et al.*, 1982). The total cost associated with these losses was estimated to be £15.2 million for the 1974 season (Jeffcott *et al.*, 1982). Increasingly, concerns are being raised about the animal welfare implications of injuries sustained by Thoroughbreds, particularly associated with the training and racing of two-year-olds and race-day fatalities.

From humble beginnings, the racing industry in Australia has developed into a large and important component of the nation’s economy. Approximately 120,000 Thoroughbreds were officially registered in 1989 (Pilkington and Wilson, 1993) and Australia produces the second highest number of Thoroughbred foals in the world after North America (ACIL, 1992). Estimates of the economic contribution of Thoroughbred racing to Australia have varied. One of the highest and widely quoted values attributed to racing rates the industry as Australia’s third largest, having a turnover of \$20 billion and employing 250,000 people¹. A more comprehensive approach was taken in a study on the Australian racing industry for State Racing Ministers (ACIL, 1992) in which the racing industry was divided into nine main activities: administration, breeding, owning, training, riding, veterinary, farriers, clubs and race gambling. It was estimated the three codes contributed 0.5% to Australia’s Gross Domestic Product (\$2,400 million) and directly employed 132,000 people in 1990/91, equating to 40,000 full-time equivalents (FTE) (ACIL, 1992). Of this, Thoroughbreds contributed \$1,806 million and 26,677 FTE (ACIL, 1992). In terms of contribution to Gross Domestic Product, the report found that the racing industry was of similar size to the agricultural output of the high rainfall zone and the iron and steel smelting industry in 1990/91 (ACIL, 1992).

Despite the size of the Thoroughbred industry in Australia and the unique features of stabling, training and racing in this country, there has been limited epidemiological research on the extent of wastage occurring here. The information thought to be of most relevance to the racing industry, yet which is lacking in the Australian literature, is the identification of risk factors for racing injuries, and the quantification of the time lost in training due to specific types of injury and disease.

IDENTIFICATION OF RISK FACTORS FOR MUSCULOSKELETAL RACING INJURIES

Racetrack injuries have a significant negative impact on the issues of animal welfare, jockey safety and the public’s perception of racing. While a number of studies investigating racetrack injuries have been conducted in countries such as the United Kingdom, the United States and Japan, results from these may not be directly applicable to the unique conditions of Australian racing. The case-control study is a type of observational study that is particularly suited to identifying risk factors for diseases that occur very infrequently (Kleinbaum,

¹ Reported in a speech delivered in 1988 by the N.S.W. Minister for Sport, Recreation & Racing (ACIL, 1992).

Kupper and Morgenstern, 1982; Schlesselman, 1982; Kelsey, Thompson and Evans, 1986), and as such, it has been used widely in the investigation of racing injuries (Mohammed, Hill and Lowe, 1991; Estberg *et al.*, 1995; Estberg *et al.*, 1996a; Kane *et al.*, 1996; Cohen *et al.*, 1997). Early case-control studies examined potential risk factors in isolation (Robinson *et al.*, 1988; Peloso, Mundy and Cohen, 1994) and did not account for confounding factors. For example, a particular racetrack may be associated with a higher risk of injury, not because of the track surface or design, but because the horses racing there are older than those at other racetracks. In this instance, age is a confounder. Matching the selection of controls to the cases with respect to certain characteristics, such as age or sex, is sometimes used in an attempt to control confounding (Estberg *et al.*, 1995; Estberg *et al.*, 1996a; Cohen *et al.*, 1997). The disadvantage is that investigation of the matched variable as a potential risk factor is not possible. The use of multivariable techniques has brought a more sophisticated approach to the investigation of racing injuries (Mohammed *et al.*, 1991; Mohammed, Hill and Lowe, 1992; Estberg *et al.*, 1996a; Kane *et al.*, 1996; Cohen *et al.*, 1997). Such techniques assess the relationship between a response variable and many explanatory variables (Thrusfield, 1995) and thus enable the effect of each putative risk factor to be determined, while controlling for all others. Potential risk factors for racing injury have included age, sex, shoe type, track, track surface, track condition, season and intensity of racing (JRA, 1991; Mohammed *et al.*, 1991; Estberg *et al.*, 1996b; Kane *et al.*, 1996).

A retrospective case-control study was undertaken to determine risk factors for serious musculoskeletal injuries ("breakdowns") in Thoroughbreds racing at the four Melbourne metropolitan racecourses during the period August 1988 - July 1995. Any horse that was recorded in the Steward's reports by the attending veterinarian as having a musculoskeletal injury and that then was either euthanised at the track or failed to race for six months from the date of injury constituted a case. Only those injuries that were clinically evident to the attending veterinarian during or immediately after the race were recorded. An equal number of control animals was selected from horses listed in the Australian Race Results (RSB, 1989 - 1995) as having raced at one of the four study tracks during the period of study without signs of injury. Variables that were thought to be potential risk factors associated with breakdown were then investigated. These included age, sex, total number of starts, days since previous start, racetrack (Flemington, Moonee Valley, Caulfield, Sandown), track condition (fast, good, dead, slow or heavy), type of race (flat, hurdle or steeplechase), class of race (stakes race or non-stakes), distance of race, field size, weight carried, barrier position, distance of previous race and change in distance from previous race, season (summer, autumn, winter or spring) and average inter-race period. An initial screening was performed to identify those variables that had little or no association with disease. Student's t test was used for continuous variables and chi-squared tests for categorical variables. All variables significant at $P \leq 0.25$ were considered eligible for inclusion in the multiple logistic regression analysis (Hosmer and Lemeshow, 1989). Backward elimination was used to determine which factors could be dropped from the multivariable model. The level of significance for a variable or factor to remain in the final model was set at 10%.

There were 196 cases of serious musculoskeletal injury in flat racing, 52 cases in hurdle races and 53 cases in steeplechases over the seven-year period. Two denominators have been used to calculate the frequency of injuries and fatalities in horse races: a *starter* refers to any horse that starts in at least one race during a racing season (the racing population); a *start* represents a single horse leaving the starting gate (race entrants) (Estberg *et al.*, 1996b; Wilson and Robinson, 1996). Injury rates can be expressed as the number of injuries per start/starter, or per 1,000 starts/starters, or as a percentage of total starts/starters. The use of starts as a risk unit is preferable when comparing risk of injury among certain race- and horse-related characteristics because it is more representative of the opportunity for injury (Estberg *et al.*, 1996b). The disadvantage of using the number of starts as a denominator is that independence of risk units may not be maintained. An example of nonindependence is when the risk of racing injury is greater for a horse that has recently raced, compared with that for its first race

of the season (Estberg *et al.*, 1996b). Although risk estimates calculated with race entrants are not true incidence rates that should use horse-time at risk as the denominator, they do provide a useful measure of the frequency of injuries and fatalities during racing (Estberg *et al.*, 1996b). Details of the incidence rates per start of cases and deaths for the three types of races are shown in Table 1.

Table 1. Incidence rates per start of cases of musculoskeletal breakdown and fatal musculoskeletal injuries (FMI) in flat, hurdle and steeplechase races at four Melbourne racecourses

Race type	Incidence of cases	Incidence of FMI
Flat	0.29% (2.9/1000 starts)	0.06% (0.6/1000 starts)
Hurdle	1.73% (17.3/1000 starts)	0.63% (6.3/1000 starts)
Steeplechase	2.91% (29.1/1000 starts)	1.43% (14.3/1000 starts)

The variables found to be the most significant risk factors for serious injury were track condition, age of the horse, racetrack and race type (Table 2). Logistic regression identified harder track surfaces, horses being older than three years of age, one of the racecourses (Flemington) and jumping races as risk factors that increased the risk of musculoskeletal breakdown.

Table 2. Risk factors for musculoskeletal breakdown in horses racing at four Melbourne racecourses

Variable	Adjusted odds ratio (95% CI)
Track condition	
Heavy	1
Slow	1.8 (0.8, 4.1)
Dead	1.3 (0.6, 2.8)
Good	2.5 (1.3, 5.1)
Fast	3.4 (1.6, 7.2)
Age	
2-3 years	1
4-5 years	1.5 (1.0, 2.3)
6-12 years	2.3 (1.4, 3.8)
Racetrack	
Moonee Valley	1
Flemington	1.9 (1.1, 3.1)
Caulfield	1.5 (0.9, 2.5)
Sandown	1.1 (0.7, 1.9)
Race type	
Flat	1
Hurdle	4.1 (2.1, 7.9)
Steeplechase	

7.9 (3.5, 17.8)

Based on the above risk factors, strategies of a practical nature that could be implemented to reduce the incidence of injury may include closer monitoring and regulation of track moisture content to avoid excessively hard racing surfaces; more rigorous examination of horses before races for signs of lameness, particularly in older horses; and altering the number and design of jumps in hurdle and steeplechase races.

USE OF SURVIVAL ANALYSIS FOR THE STUDY OF RACING CAREERS

The majority of owners and trainers seek to have their horses racing as early and for as long as possible. Most studies of racing career profiles have been descriptive, and objective information on factors that influence the time until a Thoroughbred has its first race start and the subsequent length of its racing career is limited. Survival analysis involves a response variable that measures the time until an event, being particularly useful for accommodating incomplete or censored data (Pagano and Gauvreau, 1993). It therefore has application in the study of temporal features of the racing careers of horses. In the present study, survival analysis techniques were used to evaluate factors that may influence the time until the first race start and the subsequent length of racing careers for 553 Thoroughbred horses (259 females; 294 males) catalogued for sale at the 1991 *Australian Easter Yearling Sale*.

From a commercial database that contained racing information on all horses that had made a start in Australia and New Zealand (AAP Racing Services), all racing records up to the end of the cohort's sixth year (31st July 1996) were downloaded and imported into a customised database (Microsoft Access 2.0). To this was added date of birth, obtained from the Sale Catalogue. The dates of birth were used to group horses into one of three foaling periods: early (August, September), mid (October), or late (November, December). For gender, the animal status at the time of download was recorded and subsequently categorised as male or female because the records did not indicate when colts had been castrated.

Several survival models were constructed in order to test a number of hypotheses relating to the length of time taken for a horse to make its first race start, or the length of the horse's racing career. Survival time until first race start was defined as the number of days from the first possible race start for two-year-olds (21st September 1991) to the horse's first race start in Australia or New Zealand. Career length survival time was defined as the number of days from the first to the last race start in Australia or New Zealand. For the investigation of career length, horses that raced as six-year-olds were treated as censored observations to account for the possibility of them continuing to race after the records were downloaded. Their survival time was calculated as the number of days from first race start to the end of the six-year-old racing season. Kaplan-Meier product-limit survival curves (Kaplan and Meier, 1958) were constructed to evaluate the effects of gender and foaling period on the time to first start, and the effects of gender and age at first start (two-, three-, or four-years and older) on career length. The Tarone-Ware statistic (Tarone and Ware, 1977) was used to identify statistically significant differences between survival curves.

Time to first race start

The date of the first race start was available for 461 horses. Comparison of the survival probabilities for the time to first race start revealed no significant differences between males and females ($P=0.84$) or between early-, mid- or late-born foals ($P=0.29$). None of the observations was censored.

Career length: 1st start to last start

Comparison of the survival curves for male and female horses revealed a difference in career length between the two genders (Fig.1), with male racehorses having a significantly ($P<0.01$) higher survival probability than females. The median career length was 461 days for females

and 1,013 days for males. The data for male horses contained a greater proportion of censored observations (45%) than the data for females (6%). Horses having their first race during the two-year-old season had significantly ($P < 0.01$) higher survival probabilities for career length than horses first racing at three-years or later (Fig. 2).

Figure 1. Product-limit survival curves for career length (1st to last start) in male and female Thoroughbreds

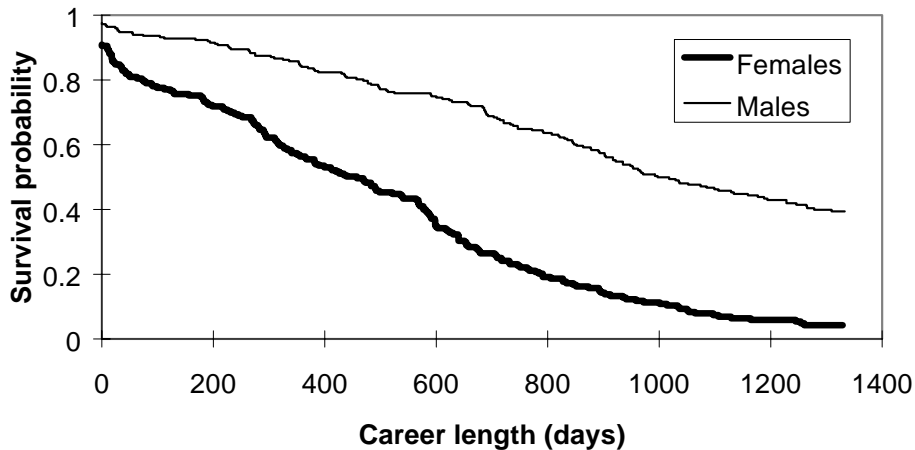
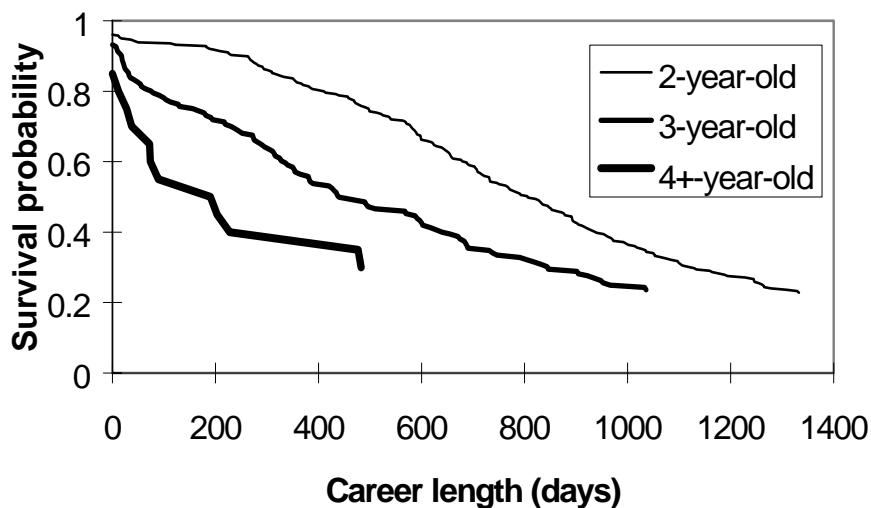


Figure 2. Product-limit survival curves for career length (1st to last start) in Thoroughbreds having their first race start at two-years of age, three-years of age or four-years of age and older



A LONGITUDINAL STUDY ON INJURIES AND DISEASE IN 2- AND 3-YEAR OLD THOROUGHBREDS IN TRAINING

Most epidemiological investigations of injuries and disease in horses in training have been prospective cohort studies, also termed follow-up or longitudinal studies. The major methodological advantage of the cohort design is that the exposure status for each subject is determined before the presence of disease is known to either the subject or to the investigator (Kleinbaum *et al.*, 1982; Rothman, 1986). Therefore, the investigator can be reasonably sure that the hypothesised cause preceded the occurrence of the disease and that the disease status did not differentially influence exposure classification (Kleinbaum *et al.*, 1982; Rothman,

1986). Other advantages of the cohort study design include the ability to calculate incidence rates in exposed and unexposed individuals, and therefore relative risk; flexibility in choosing variables to be systematically recorded; and the capability to study a range of possible health effects stemming from a single exposure (Schlesselman, 1982; Rothman, 1986; Thrusfield, 1995).

Cohort studies have the disadvantage of being statistically and practically inefficient for investigating rare diseases because information must be collected on a large number of subjects, of which only a small number will become cases (Kleinbaum *et al.*, 1982). The large numbers of subjects required to study rare diseases and the potentially long duration for follow-up, particularly if there is a long induction period between exposure and disease, make cohort studies relatively expensive to conduct (Schlesselman, 1982; Rothman, 1986; Thrusfield, 1995). Potential problems also arise from the loss of subjects because of migration, lack of participation and death. (Schlesselman, 1982; Rothman, 1986; Thrusfield, 1995).

From 525 yearlings catalogued for the 1995 *Australian Easter Yearling Sale*, 169 horses placed with participating trainers were enrolled in a longitudinal study designed to identify causes of wastage to the Thoroughbred industry. Horses were followed from the time of sale in April 1995 until the end of the cohort's three-year-old racing season, on 31st July 1997. The one investigator (CJB) personally visited the stables every 14 days and records were maintained on the training, injury and disease status of the horses in the cohort. Specifically, the following details were recorded for individual horses during each visit: weekly activity status (training in the stable, pretraining, breaking-in, or resting at pasture); presence and character of injury or disease if appropriate; and impact of such disease or injury on training. The impact on training was categorised as:

1. Days of reduced training, in which there was a reduction in training level, but not box rest or walking. For example, cantering, trotting or swimming a horse when it was scheduled to gallop.
2. Days of training prevented, in which there was a reduction in training to the level of box rest or walking.
3. Days where training was modified, the sum of 1 and 2. It is important to note that for days on which the training was modified, the horse remained in the stable and incurred full training costs.
4. Weeks rested at pasture, representing the time between the horse leaving the stable and the horse resuming either pretraining or training.

Cases of injury and disease were recorded only if the training was altered according to one of the above criteria.

The cohort horses spent 4.5% of the total observation weeks being broken in, 6.7% in pretraining, 42.4% in the stable and 46.4% resting at pasture. Of the 169 horses included in the study, 160 (95%) had entered training in the stable by the end of the cohort's two-year-old season, whereas only 76 (45%) had raced as two-year-olds. Twenty-nine horses were lost to follow-up as two-year-olds and 69 as three-year-olds, representing an attrition rate of 58%. Eighty-five per cent of horses in training suffered at least one incident of injury or disease whilst in the stable as a two-year-old. The most common injury in two-year-olds was shin soreness, which affected 42% of the 160 horses, followed by fetlock problems (25%) and coughs and nasal discharges (16%). The incidence rates for first and multiple occurrences of major injury and disease categories are presented in Table 3.

Table 3. Incidence rates for first and multiple occurrences of injury and disease in 169 Thoroughbreds observed from April 1995 to July 1997

Injury or disease category	Incidence rate for first occurrence (per 100 horse-weeks)	Incidence rate for multiple occurrences (per 100 horse-weeks)
Shin soreness	1.68	1.63
Fetlock problems	0.94	1.15
Cough / nasal discharge	0.77	0.75
Laceration / traumatic injury	0.75	0.75
Foot problems	0.32	0.36
Carpal problems	0.25	0.40

For this cohort, the total number of training days in which training was reduced (484) or prevented (591) was 1,075, representing 2.7% of the available training days. The proportion of available training days modified for two-year-olds (3.1%) was higher than for three-year-olds (2.2%). Table 4 contains the details of the days lost from training for some of the specific injury and disease categories during the study period. Lameness, excluding lacerations and traumatic injuries, was the most important cause of lost training days during the study period (56.2% of total days modified), followed by respiratory conditions (15.8%).

Table 4. Impact of injury and disease on training days in 169 Thoroughbreds observed from April 1995 to July 1997: days reduced or prevented. Conditions are listed in decreasing order, according to the percentage of total days modified

Injury / disease	Days reduced: % of 484	Days prevented: % of 591	Days modified: % of 1,075	Median days modified per case (range)
Laceration / traumatic injury	6.8%	25.0%	16.8%	4 (1 - 18)
Cough / nasal discharge	5.4%	24.4%	15.8%	5 (1 - 14)
Shin soreness	27.5%	4.4%	14.8%	4 (1 - 20)
Carpal problems	12.6%	4.6%	8.2%	3 (1 - 34)
Fetlock problems	9.3%	6.4%	7.7%	3 (1 - 10)

A total of 3,186 weeks was spent resting at pasture as a result of an injury or disease sustained in the stable, representing 20% of the total number of weeks of observation for the cohort. This was comprised of 2,350 weeks resulting from injuries in two-year-olds and 836 weeks resulting from injuries in three-year-olds. Injuries and disease sustained as three-year-olds resulted in proportionally less time resting at pasture (15.2% of three-year-old weeks followed) compared to injuries and disease sustained as two-year-olds (20.0% of two-year-old weeks followed). Table 5 contains the details of the number of weeks spent resting at pasture for some of the specific injury and disease categories during the study period. Lameness, excluding lacerations and traumatic injuries, was the most important veterinary reason for resting horses at pasture during the study period (81.2% of total weeks rested for injury or disease), followed by respiratory conditions, which included infectious and non-infectious causes (10.9%).

Table 5. Impact of injury and disease on training in 169 Thoroughbreds observed from April 1995 to July 1997: weeks rested at pasture. Conditions are listed in decreasing order, according to the percentage of total weeks rested

Injury / disease	Weeks rested at pasture: % of 3,331 ^a	Median weeks rested at pasture per case spelled (range)
Fetlock problems	23.1%	11 (1 - 44)
Shin soreness	22.5%	7 (2 - 22)
Carpal problems	8.4%	16 (2 - 31)
Cough / nasal discharge	7.9%	6.5 (2 - 30)
Miscellaneous lameness	5.6%	13 (5 - 29)

^a The total number of weeks rested at pasture in this table is greater than that listed above (3,186 weeks) because there were nine cases in which a horse was spelled for two reasons, resulting in the number of weeks rested being listed for both injuries

Although two-year-old racing in Australia is highly lucrative in terms of prize money, only 50% of elite horses race during this year. The current study has indicated that the major reason for this low figure is the high number of minor incidents that occur during the training of two-year-olds. These minor problems often alter training or result in the horse being rested, but do not prevent the horse from racing in subsequent seasons. Of the individual categories of injury or disease, lacerations and traumatic injuries, infectious respiratory disease, shin soreness, carpal problems and fetlock problems were the five most important causes of modified training days, but the impact of these on training was very different. Shin soreness, carpal problems and fetlock problems caused many days in which training was reduced but few where training was completely prevented. In contrast, infectious respiratory disease and traumatic injuries, while causing a similar number of days modified to shin soreness, had their major effect by preventing training from occurring at all. The most important individual categories of injury and disease in terms of weeks rested at pasture for the present cohort were fetlock problems, shin soreness, carpal problems and infectious respiratory disease. Therefore, whilst major injury is relatively uncommon in young horses in training, low-grade injury and disease have the potential to disrupt training schedules and cause significant economic loss.

REFERENCES

- ACIL. (1992). The contribution of the racing industry to the economy of Australia. ACIL Australia Pty Ltd in association with Cox Inall Communications Pty Ltd, Bureau of Rural Resources, Centre for Regional Economic Analysis: Canberra.
- Cohen N.D., Peloso J.G., Mundy G.D., Fisher M., Holland R.E., Little T.V., Misheff M., Watkins J.P., Honnas C.M. and Moyer W. (1997). Racing-related factors and results of prerace physical inspection and their association with musculoskeletal injuries incurred in Thoroughbreds during races. *J. Am. Vet. Med. Assoc.*, **211**, 454-463.
- Estberg L., Gardner I.A., Stover S.M., Johnson B.J., Case J.T. and Ardans A.A. (1995). Cumulative racing-speed exercise distance cluster as a risk factor for fatal musculoskeletal injury in Thoroughbred racehorses in California. *Prev. Vet. Med.*, **24**, 253-263.
- Estberg L., Stover S.M., Gardner I.A., Drake C.M., Johnson B., Case J.T. and Ardans A. (1996a). High-speed exercise history and catastrophic racing fracture in Thoroughbreds. *Am. J. Vet. Res.*, **57**, 1549-1555.
- Estberg L., Stover S.M., Gardner I.A., Johnson B.J., Case J.T., Ardans A.A., Read D.H., Anderson M.L., Barr B.C., Daft B.M., Kinde H., Moore J., Stoltz J. and Woods L.W. (1996b). Fatal musculoskeletal injuries incurred during racing and training in Thoroughbreds. *J. Am. Vet. Med. Assoc.*, **208**, 92-96.
- Jeffcott L.B., Rosedale P.D., Freestone J., Frank C.J. and Towers-Clark P.F. (1982). An assessment of wastage in Thoroughbred racing from conception to 4 years of age. *Equine vet. J.*, **14**, 185-198.
- JRA. (1991). Preventing accidents to racehorses: studies and measures taken by the Japan Racing Association. Report of the committee on the prevention of accidents to racehorses, Japan Racing Association.
- Kane A.J., Stover S.M., Gardner I.A., Case J.T., Johnson B.J., Read D.H. and Ardans A.A. (1996). Horseshoe characteristics as possible risk factors for fatal musculoskeletal injury of Thoroughbred racehorses. *Am. J. Vet. Res.*, **57**, 1147-1152.
- Kaplan E.L. and Meier P. (1958). Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.*, **53**, 457-481.
- Kelsey J.L., Thompson W.D. and Evans A.S. (1986). *Methods in Observational Epidemiology*. Oxford University Press: New York.
- Kleinbaum D.G., Kupper L.L. and Morgenstern H. (1982). *Epidemiologic Research. Principles and Quantitative Methods*. Van Nostrand Reinhold: New York.
- Mohammed H.O., Hill T. and Lowe J. (1991). Risk factors associated with injuries in Thoroughbred horses. *Equine vet. J.*, **23**, 445-448.
- Mohammed H.O., Hill T. and Lowe J. (1992). The risk of severity of limb injuries in racing Thoroughbred horses. *Cornell Vet.*, **82**, 331-341.
- Pagano M. and Gauvreau K. (1993). *Principles of biostatistics*. Duxbury Press: Belmont, California.
- Peloso J.G., Mundy G.D. and Cohen N.D. (1994). Prevalence of, and factors associated with, musculoskeletal racing injuries of Thoroughbreds. *J. Am. Vet. Med. Assoc.*, **204**, 620-626.
- Pilkington M. and Wilson G. (1993). Australian horses as a primary industry: numbers, organisation and research needs. Bureau of Resource Sciences: Canberra.
- Robinson R.A., Kobluk C., Clanton C., Martin F., Gordon B., Ames T., Trent M. and Ruth G. (1988). Epidemiological studies of musculoskeletal racing and training injuries in Thoroughbred horses, Minnesota, USA. *Acta. Vet. Scand., Suppl.*, **84**, 340-343.

- Rothman K.J. (1986). *Modern Epidemiology*. Little, Brown and Company: Boston / Toronto.
- RSB. (1989 - 1995). *Australian Race Results*. Racing Services Bureau: Melbourne.
- RSB. (1993). *Australian Thoroughbred Statistical Record*. Racing Services Bureau: Melbourne.
- Schlesselman J.J. (1982). *Case-Control Studies: Design, Conduct, Analysis*. Oxford University Press: New York.
- Tarone R.E. and Ware J. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, **64**, 156-160.
- Thrusfield M. (1995). *Veterinary Epidemiology*. Blackwell Science Ltd: Oxford.
- Wilson J.H. and Robinson R.A. (1996). Risk factors for equine racing injuries. *Compendium on Continuing Education for the Practicing Veterinarian*, **18**, 682-690.

A LONGITUDINAL STUDY OF AUSTRALIAN RACING THOROUGHBREDS: PERFORMANCE DURING THE FIRST YEARS OF RACING

Simon More

SUMMARY

A prospective longitudinal study of racing thoroughbreds is being undertaken to develop a profile of the racing careers of thoroughbred horses in south eastern Queensland, and to examine factors that affect the racing careers of these animals. Data collection for this study, which is utilising racing records about a defined cohort of horses born during 1991, commenced in 1996 and will continue until all horses cease racing. By July 31 1997, 1804 horses were enrolled in the study. This paper focuses on the performance – measured as race earnings during the first year of racing and cumulative proportion of horses still racing up to 2 years after their first start - of these horses during their first years of racing. The median race earnings during the first year of racing was A\$450, with 710 (39.4%) earning no money at all. In comparison to poor performing horses, high performing horses were more likely to be male, to have started as two year olds and to have had more starts during this year. Of the horses that first started as two and three year olds, only 46% continued racing for at least two years after their first start. Length of racing life was associated with performance during the first year of racing (number of starts and average earnings per race), and with sex, birthdate and age at first start.

INTRODUCTION

Success within the thoroughbred industry both in Australia and elsewhere is measured in terms of racing performance. Therefore, it is understandable, that intense efforts are made at each point in this industry to maximise the racing performance of individual horses.

Poor or unsatisfactory racing performance is common, and owners and trainers frequently seek professional and other assistance to correct this problem. In recent years, veterinarians have developed an increasing understanding of the causes of poor performance in individual horses (Divers and Dreyfuss, 1990), including several (such as injury and inadequate preparation) that can be remedied.

The clinical approach has proved an effective method to identify, and (in some cases) correct, the causes of poor performance in individual horses. However, it is less effective as a method to more broadly identify the factors most closely associated with high levels of racing performance in horses. This level of understanding can only realistically be gained by examining the performance of relatively large populations of horses under racing conditions. Population-based research has been used to examine the heritability of racing performance (Tolley et al., 1985), and it is now well recognised that many measures of racing performance (particularly race speed and rankings) are moderately heritable. There is as yet a much poorer understanding of non-genetic factors that affect racehorse performance, although factors that have been considered include sex, age, class of race, track condition, handicap weight and distance (Hintz 1980). An understanding of the role of these and other factors under Australian conditions may assist trainers and others to optimise the performance of the horses under their care.

Population-based research can also be used to gain an understanding of the 'normal' racing careers of thoroughbred horse populations. One of the few published studies of this type has been work undertaken with Canadian standardbreds (Physick-Sheard, 1986a,b; Physick-Sheard and Russell, 1986). This work has provided a valuable profile of the racing careers of these horses, including an objective insight into key measures being attained by horses in this industry, including levels of performance, length of racing life, and attrition rates.

The current study is being conducted using historic racing data to develop a profile of the racing careers of Australian thoroughbred horses, and to examine factors that affect the racing careers of these animals.

MATERIALS AND METHODS

General study design

This longitudinal study, with both retrospective and prospective components, is being conducted using observational epidemiological methods. The reference population is considered to be the racing thoroughbred population in Australia. In this study, the horse is the unit of interest.

Enrolment of the study cohort

Racing thoroughbreds, that met the following criteria, have been enrolled into the study:

- registered with the Racing Services Bureau (RSB; Racing Victoria Centre, Flemington)
- born on or within 12 months following 1 August 1991
- raced at least once prior to 1 August 1997, with at least one of the races during their first season of racing being at an official racetrack in the south eastern Queensland region (namely, Bundaberg, Burrandowan, Beaudesert, Caloundra, Doomben, Eagle Farm, Esk, Gatton, Gayndah, Gold Coast, Gladstone, Gympie, Ipswich, Kilcoy, Kumbia, Mt Perry, Nanango and Wondai).

Therefore, of all the RSB-registered horses born on or within 12 months of 1 August 1991, the cohort includes all those that first raced as a two year old in the 1993/94 racing season (subsequently referred to as the two year old racing cohort), as three year olds in 1994/95, as four year olds in 1995/96 and as five year olds in 1996/97, provided in each case that at least one race in their first year of racing was at an official south eastern Queensland track. Therefore, at the end of the 1996/97 racing season, four separate racing cohorts were under investigation.

These horses, subsequently considered the 'study cohort', are being observed during their full racing career.

Data retrieval

Horse and racing-level data were retrieved on a yearly basis from the RSB for all members of the study cohort. The dataset is currently complete to July 31 1997. For this study, the horse-level data includes the name of the horse; its date of birth; its country of birth; its colour and sex; the names of its sire, dam and the dam of the sire; and the name of its trainer. Note that these data reflected information supplied to the RSB by the owner, and some data – including sex and trainer – may have changed subsequently. The racing-level data related to all official racing results for each enrolled horse. For this study, these data related to the race (venue; date of race; condition of the track; distance of race; race number; number of starters; prize money; minimum weight carried; weight carried by winner) and the enrolled horse (starting position; starting odds; weight carried; whether horse was favourite; whether the horse won, race earnings).

Data management

The study records have been managed using The SAS System for Windows release 6.12 (SAS Institute Inc., Cary, NC, USA), BMDP DYNAMIC release 7 (BMDP Statistical Software Inc., Los Angeles, CA, USA) and Statistix version 4.0 (Analytical Software, Tallahassee, FL, USA).

Data analysis

Data analyses were conducted using SAS, BMDP, Statistix and Egret version 1 (Cytel Statistical Corporation, Cambridge, MA, USA). Initially, the data were examined, using frequency distributions of the original and aggregated datasets, to assess their completeness and validity. When appropriate, missing or unusual values were compared with data from other sources, including online information from the Australian Jockey Club (<http://www.ajc.org.au>) and the Australian Stud Book (<http://studbook.aust.com>). The horse and racing-level datasets were then combined to calculate aggregate information about each horse, including number of races, dates of first and last races, and total race earnings. Separate aggregate files were developed for each year of racing. Recognised methods were then used to describe these data.

Univariable and multivariable analytical methods were conducted to separately examine the association between a range of independent variables and two horse-level measures of racing performance - earnings during the first year of racing and length of racing career.

In these calculations, earnings during the first year of racing did not include money received from racing bonuses. Horses were placed into performance categories based on their first year earnings: poor and high performers earned no more than A\$450 (the median level for the cohort) and at least A\$4150 (the 75th percentile) during their first year of racing. The remaining 444 horses, that earned greater than A\$450 but less than A\$4150 during this period, were excluded from these analyses. Intra-cluster correlations, to assess whether horses with the same trainer, sire and sire of their dam were more alike than the general study cohort in terms of this performance measure, were evaluated using methods described by Donald and Donner (1988) and Donner (1993). Univariable statistics were used to determine the unconditional association between these performance categories and a range of independent variables. Variables associated at $P < 0.40$ were investigated further using a random-effect logistic regression model. Initially, a series of ordinary logistic regression models (which collectively represented each of the combinations of selected variables and their interaction terms) were constructed. The best fitting of these models was then identified, using guidelines recommended by Collett (1991), after assessing the subsequent analysis of deviance table. To account for clustering, the final model was then extended with the inclusion of the appropriate random effect variable.

Length of racing career was examined using standard methods of survival analysis (Collett, 1994). Because the dataset is not yet complete, these analyses were restricted to racing events during the two years following the first start of all horses that first raced as two and three year olds. In these calculations, horses with at least one racing event two or more years after their first start were right censored at two years. Horses without any record of racing at least two years after their first start were handled according to the date of their last recorded race. The horses that last raced prior to 1 February 1997 were considered to have ceased racing at the time of the last race. The horses that last raced after 1 February 1997 were considered to have continued racing until the end of the period of observation; in these analyses they were right censored at two years after their first start. The Kaplan-Meier method was used to calculate the cumulative proportion of horses that continued to race for up to two years after their first start. The unconditional association between this measure and a range of independent variables was assessed by calculating the log-rank (Mantel-Cox) statistic. Variables associated at $P < 0.40$ were investigated further using a proportional hazards regression model. Variables were selected in a forward stepwise manner, and retained if P remained at less than 0.05. The assumption of proportional hazards was tested by examining the effect of a $\log_e(-\log_e)$ transformation upon the survivorship function of all levels of each selected variable.

RESULTS

There are currently 1804 horses enrolled in the study, including 916 horses that first raced as two year olds, 701 as three year olds, 152 as four year olds and 35 as five year olds. Descriptive information about these horses, including comparative data from the four racing cohorts, is presented in Table 1. Within the cohort, 56.2% were males and 94.1% were born in Australia. There were significant differences between racing cohorts in terms of date of birth (the median date of birth increased with increasing age at first start) and origin of the sire (the percentage of horses with overseas sires decreased with increasing age at first start). Half of the study cohort was born on or before October 4 1991 (Figure 1).

Descriptive information about the racing performance of the horses is presented in Table 2. These horses collectively raced in 31,065 races whilst under observation, including 8661, 10848, 8282 and 3274 races during the first to fourth years of racing, respectively. During the first year of racing, horses had a median 4 starts, with 50% earning no more than a mean A\$107.60 per start. Two year olds tended to have fewer starts than other racing cohorts. In the second and subsequent years of racing, the horses tended to have more starts and greater average earnings per start in comparison to the first year of racing. The proportion of horses in the study cohort that won at least one race increased from 31.8% during the first year of racing to 54.1, 54.7 and 51.4% during the second to fourth years of racing, respectively.

The distribution of total race earnings during the first year of racing is presented in Figure 2. The data are markedly positively skewed (skewness=16.6, kurtosis=390.3). Although the median race earnings during this period was A\$450 (mean 6332.78, range 0 to 871,030), 710 (39.4%) horses earned no money at all. Based on this measure of performance, the poor performing category (those earning no more than A\$450) included 908 horses, and the high performing category (at least A\$4150) 452 horses. Unconditional associations between these categories of performance and a range of independent variables are presented in Table 3. Based on calculations of intracluster correlations for performance (measured as race earnings during the first year of racing) in data from the full study cohort, there is clear evidence of clustering at the level of the trainer and (to a much lesser extent) the sire (Table 4). Risk factors conditionally associated with performance, after accounting for trainer-level clustering, are presented in Table 5. With each additional start during the first year of racing, horses were 2.5 times more likely to be in the high rather than the poor performing category. The age of first start (2 year olds in comparison to all other racing cohorts) and sex (males in comparison to females) were each significantly associated with increased likelihood of high performance during the first year of racing.

The cumulative proportions of horses that continued to race 6 months, 1 year and 2 years following their first start were 0.83 (95% confidence interval 0.81-0.85), 0.71 (0.69-0.73) and 0.46 (0.43-0.48), respectively (Figure 3). Unconditional associations between a number of independent variables and length of racing life are presented in Table 6. Risk factors conditionally associated with cessation of racing during the 2 years following first start are presented in Table 7. Performance during the first year of racing (measured by number of starts and earnings per start) was a significant predictor of length of racing life. After accounting for all other variables in the model, three year olds were nearly twice as likely as two year olds to stop racing during the two years following first start. Sex (males in comparison to females) and date of birth (after October 1 in comparison to October 1 1991 or earlier) were also significantly associated with length of racing life in this study cohort.

Figure 1 Birthdates of horses within the study cohort

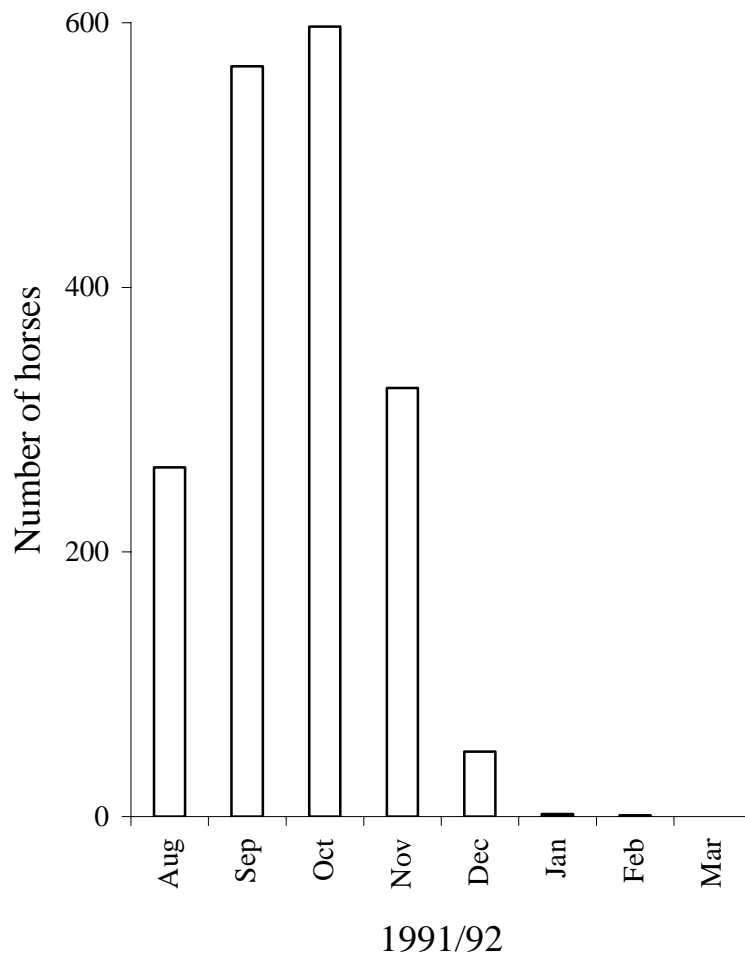


Figure 2 Total race earnings (excluding bonuses) for 1094 horses during their first year of racing. A further 710 (39.4%) horses earned no money from racing during this period

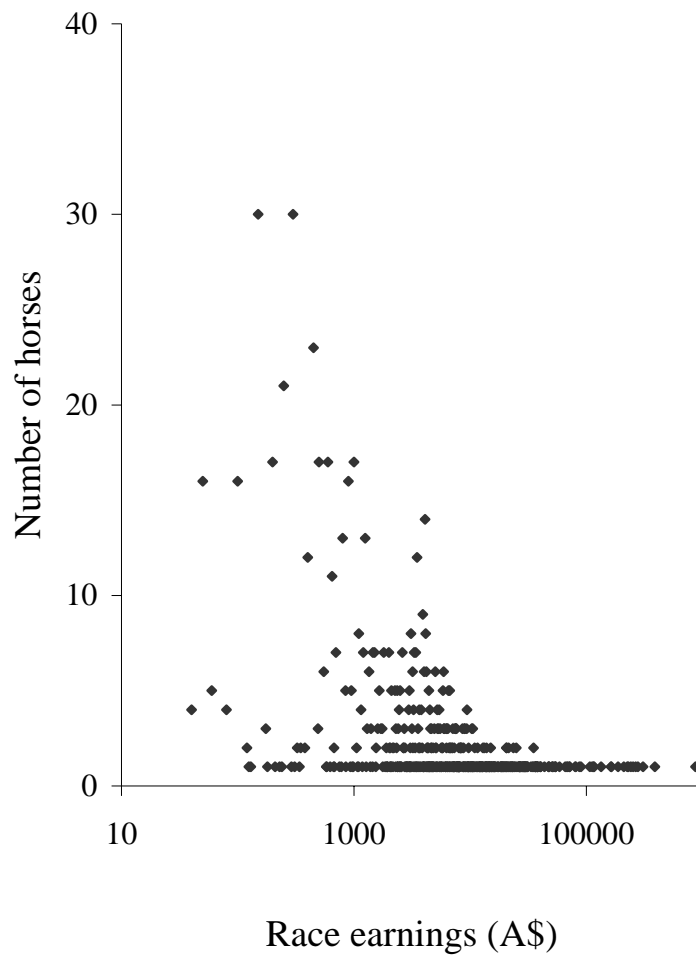


Figure 3 Cumulative proportion of 1617 enrolled horses that continued to race during the two years following their first start. The figure includes the 916 and 701 horses that first raced as 2 and 3 year olds, respectively

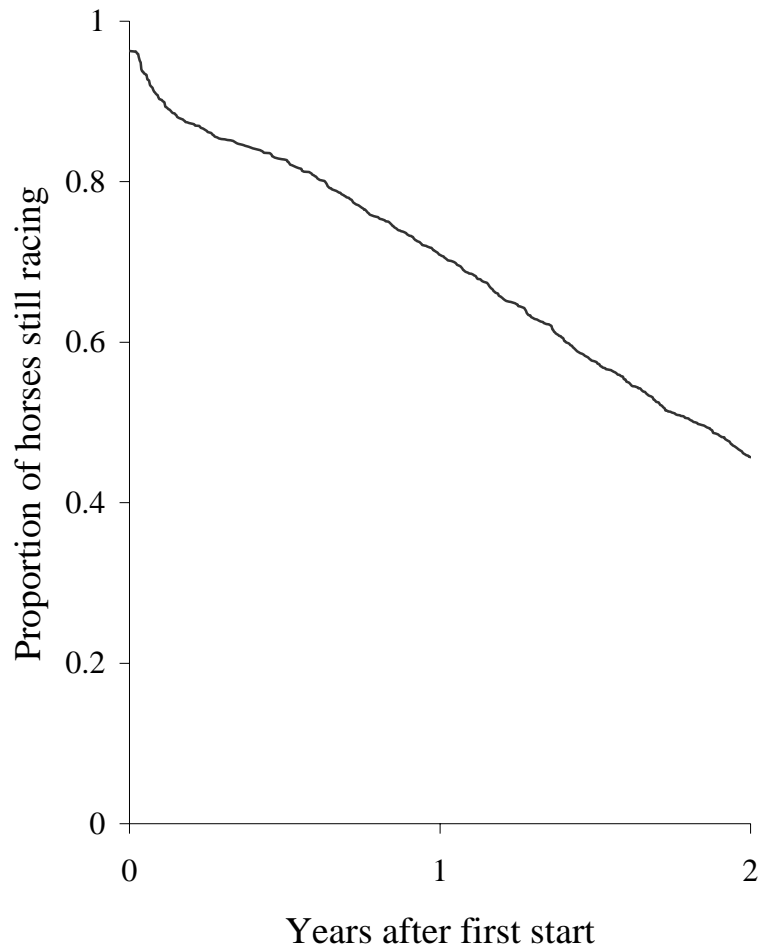


Table 1 Descriptive information about 1804 horses enrolled in the longitudinal study

Variable	All horses		Horses that first raced as:								P ^a
			2 year olds (n=916)		3 year olds (n=701)		4 year olds (n=152)		5 year olds (n=35)		
	%	Mean (Median)	%	Mean (Median)	%	Mean (Median)	%	Mean (Median)	%	Mean (Median)	
Sex (male ^b)	56.2		56.2		55.5		58.6		60.0		0.88
Colour											0.19
Black/brown	26.4		25.0		26.7		32.3		34.3		
Bay	33.1		34.5		31.4		30.9		42.9		
Chestnut	32.8		33.5		34.0		27.6		14.3		
Grey/white	7.6		7.0		7.9		9.2		8.5		
Date of birth		5OCT91 (4OCT91)		1OCT91 (1OCT91)		7OCT91 (6OCT91)		12OCT91 (12OCT91)		14OCT91 (8OCT91)	<0.001
Born in Australia ^c	94.1		95.4		92.4		92.1		100		0.02
Sire was born in Australia ^d	41.2		36.2		45.8		46.1		57.1		<0.001
Dam was born in Australia ^e	78.4		78.2		78.6		78.3		80.0		0.99
Sire of the dam was born in Australia ^f	28.5		26.7		29.8		31.6		37.1		0.27

a Difference between horses in the four racing cohorts; P value of chi-square (for categorical data) or Kruskal-Wallis (for continuous data) statistic

b Male horses have been recorded as either geldings or stallions in the RSB database. These data are likely to accurately reflect the status of these horses at the time of recording. However, these data may change over time, and no attempt has been made to differentiate them here

c All 107 non-Australian horses were from New Zealand

d Non-Australian horses originated from USA (533 horses), New Zealand (257), Ireland (126), Great Britain (83), Canada (30) and France (32)

e Non-Australian horses originated from New Zealand (254 horses), USA (60), Ireland (35), Great Britain (28), France (8), Canada (3), Italy (1) and Argentina (1)

f Non-Australian horses originated from USA (427 horses), Great Britain (335), Ireland (208), New Zealand (128), France (125), Canada (58), Italy (4), Argentina (3) and Sweden (1)

Table 2 The racing performance of 1804 horses enrolled in the longitudinal study

Variable	All horses		Horses that first raced as:								P ^a
			2 year olds		3 year olds		4 year olds		5 year olds		
	%	Mean (Median)	%	Mean (Median)	%	Mean (Median)	%	Mean (Median)	%	Mean (Median)	
<i>During the first year of racing^b</i>											
Number of starts		4.8 (4.0)		4.19 (3.0)		5.5 (5.0)		5.4 (4.0)		5.1 (4.0)	<0.001
Average earnings per start (A\$) ^c		902.48 (107.60)		1219.41 (100.0)		628.44 (145.0)		421.86 (62.50)		183.87 (64.29)	0.08
Horses that won at least one race	31.8		28.9		36.2		30.9		22.9		0.01
<i>During the second year of racing^d</i>											
Number of starts		7.9 (7.0)		7.7 (7.0)		8.2 (8.0)		7.6 (7.0)			0.05
Average earnings per start (A\$) ^c		1090.18 (394.10)		1135.60 (400)		1109.08 (386.36)		597.73 (397)			0.75
Horses that won at least one race	54.1		54.1		54.5		51.6				0.88
<i>During the third year of racing^e</i>											
Number of starts		9.2 (8.0)		9.3 (9.0)		9.0 (8.0)					0.54
Average earnings per start (A\$) ^c		1125.57 (419.44)		1163.27 (460.0)		1058.86 (362.44)					0.04
Horses that won at least one race	54.7		57.4		50.0						0.03
<i>During the fourth year of racing^f</i>											
Number of starts		8.9 (7.0)		8.9 (7.0)							
Average earnings per start (A\$) ^c		992.81 (413.39)		992.81 (413.39)							
Horses that won at least one race	51.4		51.4								

a Difference between horses in the racing cohorts; P value of chi-square (for categorical data) or Kruskal-Wallis (for continuous data) statistic

b Data about the first year of racing are available for 1804 horses (and 8661 racing events), including 916 that first raced as 2 year olds, 701 as 3 year olds, 152 as 4 year olds and 35 as 5 year olds

c Excludes racing bonuses

d Data about the second year of racing are available for 1370 horses (& 10848 racing events), including 778 (84.9% of the) horses that first raced as 2 year olds, 501 (71.5) as 3 year olds and 91 (59.9) as 4 year olds

e Data about the third year of racing are available for 903 horses (& 8282 racing events), including 577 (63.0% of the) horses that first raced as 2 year olds and 326 (46.5) as 3 year olds

f Data about the fourth year of racing are available for 368 horses (& 3274 racing events), representing 40.2% of the horses that first raced as 2 year olds

Table 3 Association between a number of independent variables and performance during the first year of racing for 1360 enrolled poor and high performing horses, ordered according to level of statistical relationship

Variable	Poor performers ^a (n=908)		High performers ^b (n=452)		P ^c
	%	Mean (Median)	%	Mean (Median)	
Sex (stallions and geldings)	51.2		66.6		<0.001
Number of starts		2.8 (2.0)		8.0 (7.0)	<0.001
Horse originated from Australia	95.7		90.7		<0.001
Dam of horse originated from Australia	80.8		74.3		0.006
Sire of dam of horse originated from Australia	30.6		26.1		0.08
Sire of horse originated from Australia	42.5		38.5		0.16
Age at first race					0.36
Two years	52.6		51.5		
Three years	36.1		39.6		
Four years	8.9		7.5		
Five years	2.3		1.3		
Colour					0.47
Black/brown	27.1		25.0		
Bay	32.2		35.0		
Chestnut	32.6		33.6		
Grey/white	8.1		6.4		
Date of birth		5OCT91 (4OCT91)		4OCT91 (3OCT91)	0.99

a Horses that earned no more than A\$450 (the median value for the overall cohort) during the first year of racing

b Horses that earned at least A\$4150 (the 75th percentile for the overall cohort) during the first year of racing

c P value of chi-square (for categorical data) or Kruskal-Wallis (for continuous data) statistic

Table 4 Intracluster correlation for the performance (measured as race earnings during the first year of racing) of 1804 enrolled horses, clustered by sire, dam, sire of the dam and trainer

Cluster	Number of clusters	Cluster size			Intracluster correlation ^a
		Mean	Median	Range	
Sire	508	3.6	2.0	1-41	0.10
Dam	1804	1.0	1.0	1	- ^b
Sire of the dam	828	2.2	1.0	1-28	-0.29
Trainer ^c	966	1.7	1.0	1-31	0.78

a Intracluster correlation is considered low (<0.05), medium (0.05-0.10) and high (>0.10). Due to computational limitations, these correlations were calculated using 200 randomly chosen levels from each cluster

b Not calculated because the mean cluster size was 1.0

c Data about trainers were missing for 118 horses

Table 5 Risk factors conditionally associated with the performance (measured as race earnings during the first year of racing) of 1271 enrolled horses^a

Variable	Beta coefficient	Standard error (beta)	p	Adjusted odds ratio	95% CI (OR)
Intercept	-5.95	0.60			
Number of starts	0.90	0.08	<0.001	2.5 ^b	2.1-2.9
Age at first race					
3yos versus 2 yos	-1.02	0.25	<0.001	0.4 ^c	0.22-0.59
4yos versus 2 yos	-0.97	0.42	0.02	0.4	0.17-0.86
5yos versus 2yos	-2.09	1.04	0.05	0.1	0.02-0.95
Sex (males versus females)	0.75	0.22	<0.001	2.1	1.4-3.3
Trainer-level random effect intercept ^d	1.13	0.26			

Deviance of 826.5 with 1264 degrees of freedom. In the ordinary logistic regression model, deviance was 843.7 with 1265 degrees of freedom

a Although 1360 horses were assigned to the categories of poor or high performance, data about trainers were missing for 89 of these animals

b That is, after taking account of each of the other variables in the model, for each additional race run during the first year of racing, there was an increasing likelihood (of 2.5 times) that these horses were considered high rather than poor performing animals

c That is, after taking account of each of the other variables in the model, horses that first raced as three year olds were much less (0.4 times as) likely to be high performers during their first year of racing than those that first started as two year olds

d Inclusion of a trainer-level random effect resulted in a significant improvement in the model (likelihood ratio statistic of 17.3 with 1 degree of freedom, $p < 0.001$)

Table 6 Association between a number of independent variables and length of racing life for 1617 enrolled horses (which includes the 916 and 701 horses that first raced as 2 and 3 year olds, respectively), ordered according to level of statistical relationship

Variable	Number of horses		Subsequent to their first race, the proportion still racing at			P ^b
	Total	Ceased racing within 2 years ^a	6 months	1 year	2 years	
Sex						<0.001
Males	904	401	0.86	0.76	0.56	
Females	713	476	0.78	0.64	0.33	
Age at first race						<0.001
Two years old	916	441	0.88	0.77	0.52	
Three years old	701	436	0.76	0.63	0.38	
Number of starts during first year of racing ^c						<0.001
Four or less	929	577	0.73	0.62	0.38	
Greater than four	688	300	0.95	0.83	0.56	
Earnings per start during the first year of racing ^d						<0.001
No more than A\$108	801	547	0.72	0.57	0.32	
More than A\$108	816	330	0.94	0.85	0.60	
Date of birth ^e						0.06
4OCT91 or earlier	833	473	0.83	0.70	0.43	
Later than 4OCT91	784	404	0.83	0.72	0.48	
Origin of the sire						0.06
Australia	653	372	0.81	0.69	0.43	
Outside Australia	964	505	0.84	0.72	0.48	
Colour						0.19
Black/brown	416	240	0.80	0.69	0.42	
Bay	536	278	0.86	0.75	0.48	
Chestnut	545	290	0.83	0.70	0.47	
Grey/white	120	69	0.78	0.66	0.42	
Origin of the dam						0.53
Australia	1267	679	0.82	0.70	0.46	
Outside Australia	350	198	0.84	0.72	0.43	
Origin of the sire of the dam						0.64
Australia	454	250	0.82	0.70	0.45	
Outside Australia	1163	627	0.83	0.71	0.46	
Origin of horse						0.87
Australia	1522	824	0.82	0.71	0.46	
Not Australia	95	53	0.92	0.75	0.44	

a In these calculations, horses with at least one racing event two or more years after their first start were right censored at two years. Horses without any record of racing at least two years after their first start were handled according to the date of their last recorded race. The horses that last raced prior to 1 Feb 1997 were considered to have ceased racing at the time of the last race. The horses that last raced after 1 Feb 1997 were considered to have continued racing until the end of the period of observation; in these analyses they were right censored at two years after their first start

b P value of logrank (Mantel-Cox) statistic

c The full cohort raced a median 4 races during their first year of racing

d The full cohort earned a median A\$107.60 per start during their first year of racing

e The median date of birth for the full cohort was October 4 1991

Table 7 Risk factors conditionally associated with cessation of racing for 1617 enrolled horses (which includes the 916 and 701 horses that first raced as 2 and 3 year olds, respectively) during the first two years following first start

Variable	Beta coefficient	Standard error (beta)	Adjusted relative risk	95% CI (RR)
Age at first racing (3yos versus 2 yos)	0.63	0.07	1.88 ^a	1.65-2.16
Sex (males versus females)	-0.53	0.07	0.59	0.52-0.67
Date of birth (after 4Oct91 versus 4Oct91 or earlier)	-0.13	0.07	0.88	0.77-1.0
During the first year of racing:				
Number of starts (greater than 4 versus 4 or less)	-0.40	0.08	0.67	0.57-0.79
Earnings per start (greater than A\$108 versus A\$108 or less)	-0.74	0.08	0.48	0.41-0.56

a That is, after accounting for all other factors in the model, horses that first raced as three year olds were nearly twice as likely to stop racing during the first two years following their first race than horses that first raced as two year olds

DISCUSSION

Epidemiological methods are gaining increasing recognition in the study of problems of equine health and performance (Reeves and Smith, 1995). These methods have been used to issues including wastage or training failure (Bailey et al., 1997; Lindner and Dingerkus, 1993), racetrack fatalities and injuries (Estberg et al., 1995; Estberg et al, 1996; Estberg et al., 1998; McKee, 1995; Mohammed et al., 1991) and several medical conditions including colic (Cohen and Peloso, 1996; Reeves et al., 1990) and clinical cyathostomiasis (Reid et al., 1995). Epidemiological methods have also been used to develop a career profile of Canadian standardbreds (Physick-Sheard, 1986a,b; Physick-Sheard and Russell, 1986). This paper presents preliminary results from a similar longitudinal study about Australian thoroughbred horses. Utilising existing racing data, this study is being conducted to develop a profile of the racing careers of a defined cohort of Australian thoroughbreds, and to examine factors that affect the racing careers of these animals.

Following each racing season, the RSB has routinely archived the racing records of all horses that had not started during the previous two years. Consequently, the active RSB database has only ever held a complete record of racing events during the previous two racing seasons. Because it has not been possible to access archived data, data retrieval for this observational study first commenced following the completion of the 1995/96 racing season. Retrieval of data for this study will continue on a yearly basis until all horses within the cohort have ceased racing. This paper presents results from the analyses of data that was current to July 31 1997.

The selection criteria were devised to minimise the possibility of sampling bias, and it is likely that the study cohort is representative of the general Australian thoroughbred population. The RSB is the central repository for all thoroughbred racing records. Furthermore, all horses need to be registered with the RSB before starting in any official race in Australia or New Zealand. I am unaware of any systematic difference between the 1991/92 birth cohort and thoroughbreds born at other times during this general period. Although there are regional differences in aspects of the thoroughbred industry in Australia, southeastern Queensland is probably typical of the racing industry in the vicinity of any large Australian urban centre.

Over half (50.8%) of the horses in the study cohort were first started as two year olds. Although these horses did not differ in most respects from horses that were started at a later age, they were more likely to have been born earlier during the 1991/92 season, and been sired by a horse not born in Australia. In common with previous findings (Bourke, 1995), a majority of the horses were males.

The performance of the study cohort – measured as either the race earnings per start or total race earning per season – was extremely skewed. For example, during the first year of racing, the mean and median race earnings were A\$6,332.77 and A\$450. Although the highest earnings was A\$871,030, 710 (39.4%) horses earned no money at all. Based on these figures, very few owners would have been able to recoup training costs from race earnings during the first year of racing. Training costs would likely equal or excel A\$8,400 (estimated on the basis of two training preparations at A\$50/day for a combined 24 weeks during the year), whereas returns from racing exceeded this sum in only 237 (13.1%) horses within the study cohort. Note that these estimates take no account of bonuses that would have been paid to many of these horses at each start; they also ignore additional costs such as racing, track or veterinary fees.

It is noteworthy that racing performance (measured as the total race earnings during the first year of racing) is clustered at the level of the trainer and (to a much lesser extent) the sire. Therefore, horses within a single stable are much more alike, in terms of this measure of performance, than horses from other stables. This is not surprising, given the ability of a trainer's reputation or fee to attract horses of similar quality. Training methods may also have an influence on these findings. Sire-effects are also expected because of the moderate heritability of performance traits in racing thoroughbreds (Tolley et al., 1985).

For some years, there have been concerns about the adverse and long term effects of racing on horses that first start racing as two year olds. In particular, most two year old horses have evidence of skeletal immaturity (based on radiographic studies of the time of closure of the distal radial epiphyses) for at least the first part of their first season of racing (Mason and Bourke, 1973; Ranganathan and Bhasker, 1986). This condition has been linked to a range of injuries that occur commonly in two year old animals, including shin soreness and carpalitis (Mason and Bourke, 1973). Indeed, 40-80% of two year old Australian horses are believed to develop shin soreness during their first year of racing (Bourke, 1990). Although these conditions can result in significant losses due to forced periods of rest and/or treatment and lost opportunities to race, a recent study has demonstrated that two year old racing does not shorten subsequent racing life (Bourke, 1990; Bourke, 1995). The results of the current study confirm these findings. Indeed, within the study cohort, the two year old racing cohort outperformed the older racing cohorts in each of the two measures of performance under investigation. These findings were not confounded by those factors (date of birth, origin of horse, origin of sire) that were known to significantly differ between the racing cohort.

In the survival analyses, the outcome (cessation from racing) was of interest for up to two years following each horse's first start. Because the three year racing cohort could have included horses that first started as late as July 31 1995 (and therefore remained under observation until the end of the period when data were available for this paper), these analyses were restricted to those horses that started as two and three year olds. In these calculations, because no data are yet available beyond July 31 1997, it is possible that some misclassification may have occurred with respect to the date that some horses ceased to race. The horses in the two year old racing cohort were each under observation for at least one year following the two year observation period and it is very unlikely that 'inactive' horses would have subsequently recommenced racing after July 31 1997. Similarly, it is unlikely that those horses from three year old racing cohort with their most recent racing event recorded prior to February 1 1997 horses were incorrectly classified as 'inactive'. In either circumstance, it is noteworthy that the maximum average interval between starts (for horses that raced at least five races during these years) was 66.2, 68, 71.4 and 68.4 days, during the first to fourth years

of racing, respectively. However, it is possible that misclassification may have occurred with the remaining 21 (3.0%) horses from the three year old racing cohort. These animals, each with at least one racing event on or later than February 1 1997, were considered right censored at two years following their first start for the purpose of these analyses. However, it is possible that some may in fact have raced on no further occasions beyond July 31 1997. If misclassification of some of these horses has occurred, the levels of wastage will be slightly more pessimistic than presented here.

This study has confirmed a very high level of wastage (defined as the premature retirement of horses from racing) among Australian racing thoroughbreds. Of the 1617 horses within the study cohort that first raced as two or three year olds, only 71% continued to race for at least one and 46% for at least two years after their first start. In earlier work, Bourke (1995) estimated that over one-third of the Victorian racehorse population is replaced each year. Wastage in racing animals has been attributed to lack of ability and the development of exercise-induced injury (Jeffcott et al., 1982; Bourke, 1995). While it is not possible to identify the main causes of wastage in the current study, it seems likely that lack of ability is an important contributor. Performance during the first year of racing was significantly associated with the likelihood that horses would cease racing at any stage during the first two years following their first start. In this study, other predictors of racing life included age at first racing, sex and date of birth. As with earlier findings (Bourke, 1995), males and horses that first started at two years of age were less likely to cease racing during the first two years following their first start than females and horses that first started as three year olds. The effect of date of birth (late births more favourable than early births) on subsequent length of racing life has not been described previously. Indeed, these findings are in contradiction to Australian industry practice where breeders aim to have mares foal each year as close following August 1 as possible. Because the adjusted relative risk for this association could span unity 5% of the time, it is important that further work be conducted to re-examine this finding.

ACKNOWLEDGMENTS

I thank Richard Lindsey, Steven Runnalls and Thapliyal Deepak for their help with the retrieval of data from the RSB database and Rodney Verrall for helpful discussions about the Australian racing industry.

REFERENCES

- Bailey, C.J., Rose, R.J., Reid, S.W.J. and Hodgson, D.R., (1997). Wastage in the Australian thoroughbred racing industry: a survey of Sydney trainers. *Australian Veterinary Journal*, **75**: 64-66.
- Bourke, J.M., (1995). Wastage in thoroughbreds. Proceedings from the Annual Seminar Equine Branch, New Zealand Veterinary Association, Auckland, July 1995. *Veterinary Continuing Education*, Massey University, Publication No. 167.
- Bourke, J.M., (1990). Two year old racing. *Australian Equine Veterinarian*, **8**:146-147.
- Cohen, N.D. and Peloso, J.G., (1996). Risk factors for history of previous colic and for chronic, intermittent colic in a population of horses. *Journal of the American Veterinary Medical Association*, **208**: 697-703.
- Collett, D., (1994). Modelling survival data in medical research. Chapman and Hall, London.
- Collett, D., (1991). Modelling binary data. Chapman and Hall, London.
- Divers, T.J. and Dreyfuss, D., (1990). Evaluating the horse with a poor racing performance. *Veterinary Medicine*, **85**: 522-529.
- Donald, A. and Donner, A., (1988). The analysis of variance adjustment to chi-square tests when there is litter or herd correlation in surveys or field trials. *Acta Veterinaria Scandinavica*, Supplement, **84**:490-492.
- Donner, A., (1993). The comparison of proportions in the presence of litter effects. *Preventive Veterinary Medicine*, **18**:17-26.
- Estberg, L., Gardner, I.A., Stover, S.M. and Johnson, B.J., (1998). A case-crossover study of intensive racing and training schedules and risk of catastrophic musculoskeletal injury and lay-up in California thoroughbred racehorses. *Preventive Veterinary Medicine*, **33**: 159-170.
- Estberg, L., Gardner, I.A., Stover, S.M., Johnson, B.J., Case, J.T. and Ardans, A., (1995). Cumulative racing-speed exercise distance cluster as a risk factor for fatal musculoskeletal injury in thoroughbred racehorses in California. *Preventive Veterinary Medicine*, **24**: 253-263.
- Estberg, L., Stover, S.M., Gardner, I.A., Johnson, B.J., Case, J.T., Ardans, A., Read, D.H., Anderson, M.L., Barr, B.C., Daft, B.M., Kinde, H., Moore, J., Stoltz, J. and Woods, L.W., (1996). Fatal musculoskeletal injuries incurred during racing and training in thoroughbreds. *Journal of the American Veterinary Medical Association*, **208**: 92-96.
- Hintz, R.L., (1980). Genetics of performance in the horse. *Journal of Animal Science*, **51**: 582-594.
- Jeffcott, L.B., Rosedale, P.D., Freestone, J., Frank, C.J. and Towers-Clark, P.F., (1982). An assessment of wastage in thoroughbred racing from conception to 4 years of age. *Equine Veterinary Journal*, **14**: 185-198.
- Lindner, A. and Dingerkus, A., (1993). Incidence of training failure among thoroughbred horses at Cologne, Germany. *Preventive Veterinary Medicine*, **16**: 85-94.
- Mason, T.A. and Bourke, J.M., (1973). Closure of the distal radial epiphysis and its relationship to unsoundness in two year old thoroughbreds. *Australian Veterinary Journal*, **49**: 221-228.
- McKee, S.L., (1995). An update on racing fatalities in the UK. *Equine Veterinary Education*, **7**: 202-204.
- Mohammed, H.O., Hill, T. and Lowe, J., (1991). Risk factors associated with injuries in thoroughbred horses. *Equine Veterinary Journal*, **23**: 445-448.

Physick-Sheard, P.W., (1986a). Career profile of the Canadian standardbred. I. Influence of age, gait and sex upon chances of racing. *Canadian Journal of Veterinary Research*, **50**: 449-456.

Physick-Sheard, P.W., (1986b). Career profile of the Canadian standardbred. II. Influence of age, gait and sex upon number of races, money won and race times. *Canadian Journal of Veterinary Research*, **50**: 457-470.

Physick-Sheard, P.W. and Russell, M., (1986). Career profile of the Canadian standardbred. III. Influence of temporary absence from racing and season. *Canadian Journal of Veterinary Research*, **50**: 471-478.

Ranganathan, M. and Bhasker, C.G., (1986). Radiographic study of long bones in two year old race horses. *Centaur*, **2**: 113-166.

Reeves, M.J., Curtis, C.R., Salman, M.D., Reif, J.S. and Stashak, T.S., (1990). A multivariable prognostic model for equine colic patients. *Preventive Veterinary Medicine*, **9**: 241-257.

Reeves, M.J. and Smith, G., (1995). The application and use of epidemiological modelling in the study of equine health problems. *Equine Veterinary Journal*, **27**: 84-85.

Reid, S.W.J., Mair, T.S., Hillyer, M.H. and Love, S., (1995). Epidemiological risk factors associated with a diagnosis of clinical cyathostomiasis in the horse. *Equine Veterinary Journal*, **27**: 127-130.

Tolley, E.A., Notter, D.R. and Marlowe, T.J., (1985). A review of the inheritance of racing performance in horses. *Animal Breeding Abstracts*, **53**: 163-185.

APPLICATION OF EPIDEMIOLOGICAL TECHNIQUES TO STUDIES OF EQUINE DISEASE

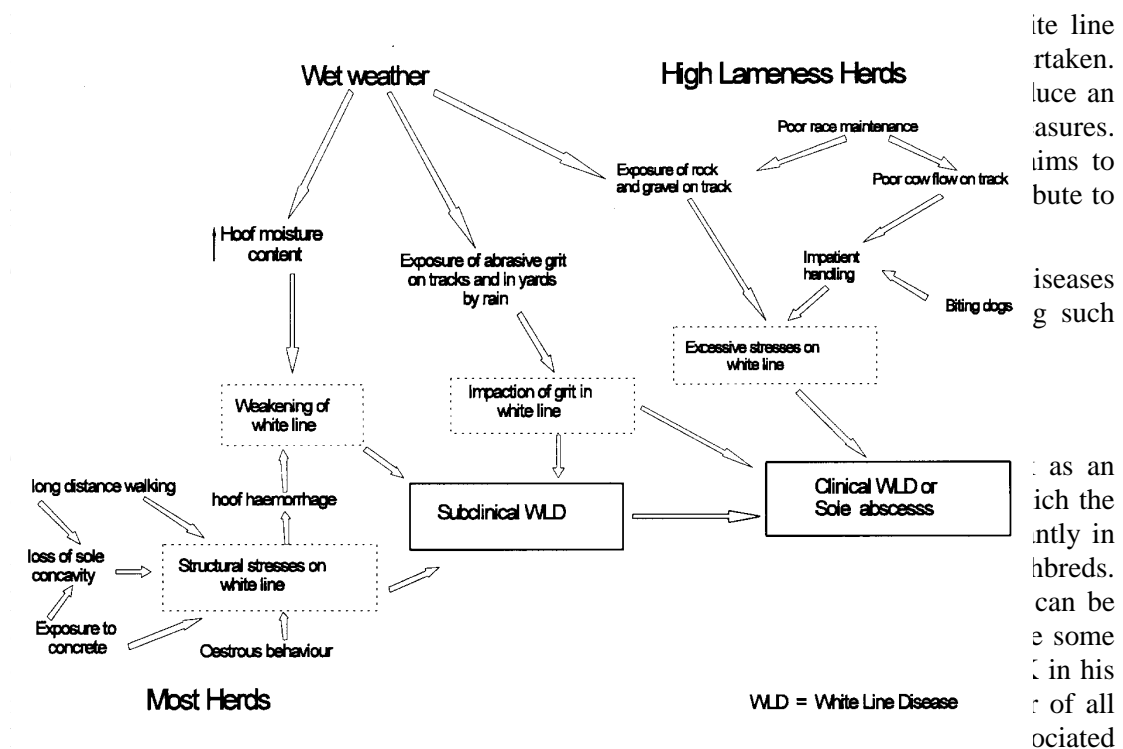
Roger Morris and Nigel Perkins

INTRODUCTION - AN OVERALL APPROACH

In this paper we will illustrate how epidemiological investigation methods can be used to examine health issues in the horse (both diseases and injuries), and the way in which newer analytical methods can contribute to extracting maximum value from such studies. Each of the major study methods will be considered in turn, from those which offer relatively weak inference to those which provide strong inference about the influence of risk factors, and examples of the use of each method will be given. Further examples for racing injuries are provided in the review by Wilson and Robinson (1996).

Views of the epidemiology of disease have evolved over time from the “Koch’s postulates” view of disease as a linear process driven by the agent, to the more realistic “host-agent-environment” triangle (in which these three factors interact to determine the disease process), to the current “causal web” or ecosystem view of disease.

Figure 1. A system diagram for one form of lameness in dairy cows (white line disease) that shows likely causal links between risk factors that predispose to both subclinical and clinical disease



with adverse outcomes can realistically be created in what is fundamentally a case series study. While inferences about all horses or all thoroughbreds in the catchment areas of the hospitals would be very weak, reasonable inferences can be drawn about the population of “horses presented for anaesthesia”. This will help the surgeon to advise wisely on the prognosis for such animals, and to take action on measures to minimise deaths associated with anaesthesia. When designed and reported in this framework, valid comparisons can be made in various ways. The data can be presented in a cohort study format (eg the cohort of horses exposed to halothane as an anaesthetic agent versus those exposed to other agents) and analyzed as a relative risk. It may also be possible within such a study to create from the data

a *nested case-control study* on issues of particular interest - for example to examine the cases of horses which broke a leg while recovering from anesthesia in comparison with random and/or matched controls which did not, to evaluate risk factors associated with such injuries using odds ratios. The case series study, if conducted in this format, may even lead on directly to an *intervention study* if evidence is strong enough that preventive measures are possible, such as Johnston's most recent comparative evaluation of halothane and isoflurane for maintenance of anaesthesia in the horse, using a randomised clinical trial format.

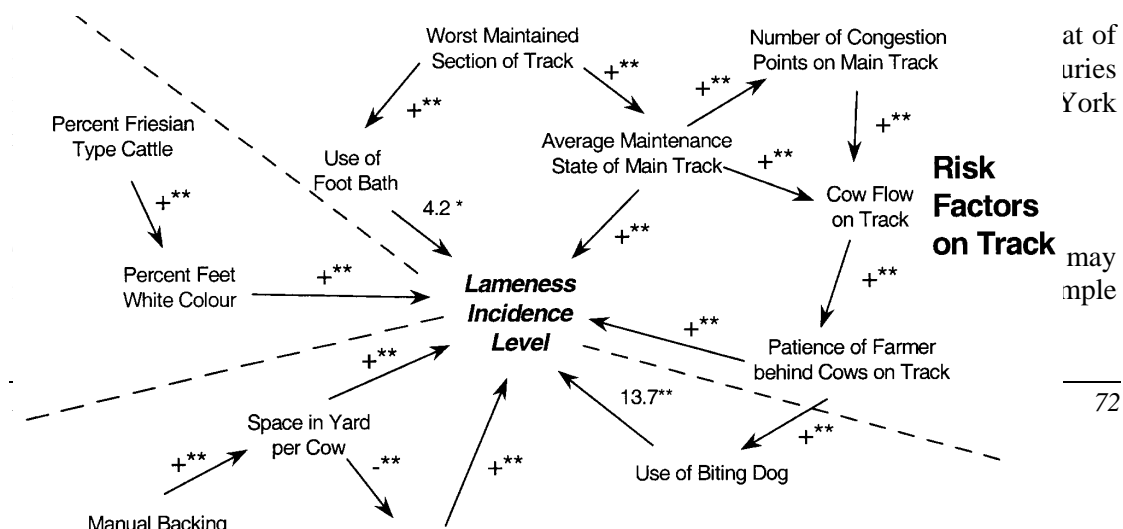
It is also possible to use accumulated case records to examine associations among factors involved in a case load at a veterinary hospital (Reeves *et al* 1989), using odds ratios (particularly in the form of cross-tabulations using the Mantel-Haenszel stratified analysis) to identify high risk groups for particular findings and outcomes. Like all studies of this general nature, it depends on adequate case definitions being used, and suffers from all the problems described above, plus variability in completeness of diagnostic procedures and adequacy of record-keeping. However for exploratory purposes, such retrospective studies can provide useful initial guides to future research directions.

CASE-CONTROL STUDY

This is the most widely used and useful exploratory technique for investigating diseases and injuries, especially for uncommon or unpredictable problems. Cases can be chosen as they become available, and controls may be random, or if necessary matched on one or two factors which the analyst wishes to prevent from influencing the interpretation of risk factor effects. In our work we have made use of both random and matched sets of controls in two separate analyses based on the same cases, since this allows us to focus the main analysis on the random controls, but examine special factors in the matched analysis which would be confounded in the random comparison.

Analysis of case-control data is usually either through cross-tabulation tables, or in more complex analyses by multiple logistic regression (regression with a binary outcome such as affected/unaffected). If the controls are random then standard logistic regression is appropriate, but for matched controls conditional logistic regression analysis is required - which is only available in a few statistical packages. Usual practice in such analyses is to screen variables using univariate analytical methods, and then to put variables which show evidence of a possible association with the outcome variable (up to about $P < 0.15$ or 0.2) into the multivariate analysis, plus carefully selected factors which on biological grounds appear to deserve consideration even though the univariate association is very weak. The final model can be produced using backward elimination or best sub-sets regression. Another very useful analytical approach in exploratory investigations of epidemiological data to develop causal webs is path analysis, sometimes called "the thinking man's regression analysis". This can be used with either a linear or logistic model, and from a starting model which represents the factors of potential importance in the disease, the analytical process removes non-significant arrows to produce the final path model containing only statistically significant relationships.

Figure 2. Example of logistic path modelling of bovine lameness – final path model



which “represents” in some form the population of interest, and allows valid inferences to be drawn about that population. Sampling animal populations is commonly complicated by the fact that they are kept in groups, and that a small proportion of the groups comprise a large proportion of the total animals. Sampling groups using random selection procedures (such as by interviewing owners or trainers) provides valid data on groups but a biased sample of animals, whereas basing the sample on animals without regard to ownership will represent the animals fairly but provide a biased group of owners (and is very challenging to conduct because it effectively requires a census of the population of animals). Providing unbiased data on both animals and owner or trainer groups is best achieved by the “probability of animal selection proportional to group size” approach, which is adequately approximated by stratified random sampling using groups size to define strata.

Such studies are useful for determining the characteristics of horse populations (Kaneene *et al* 1997a), the views of owners and managers concerning health issues (Bailey *et al* 1997; Kaneene *et al* 1996), and measuring the prevalence of various diseases (Kaneene *et al* 1997b) and of various putative (hypothesised) risk factors for the disease(s), including the proportion of animals which have both the risk factor and disease. However they tell us nothing about disease processes since they cannot measure incidence or duration of disease, which are key determinants of the importance of diseases. To do this requires repeated measurements on two or more occasions, identifying and re-sampling the same animals. However this is then better described as a longitudinal study, which is discussed below. If animals are not identified through sequential examinations then it is a repeated cross-sectional study, which has far more limited value than the true longitudinal approach since neither incidence nor duration can be confidently measured.

Cross-sectional studies are analysed by relatively simple summarising methods, since they have limited direct inferential value on causal relationships, but they can provide comprehensive information on populations, to an extent rarely achieved in other study designs.

COHORT STUDY

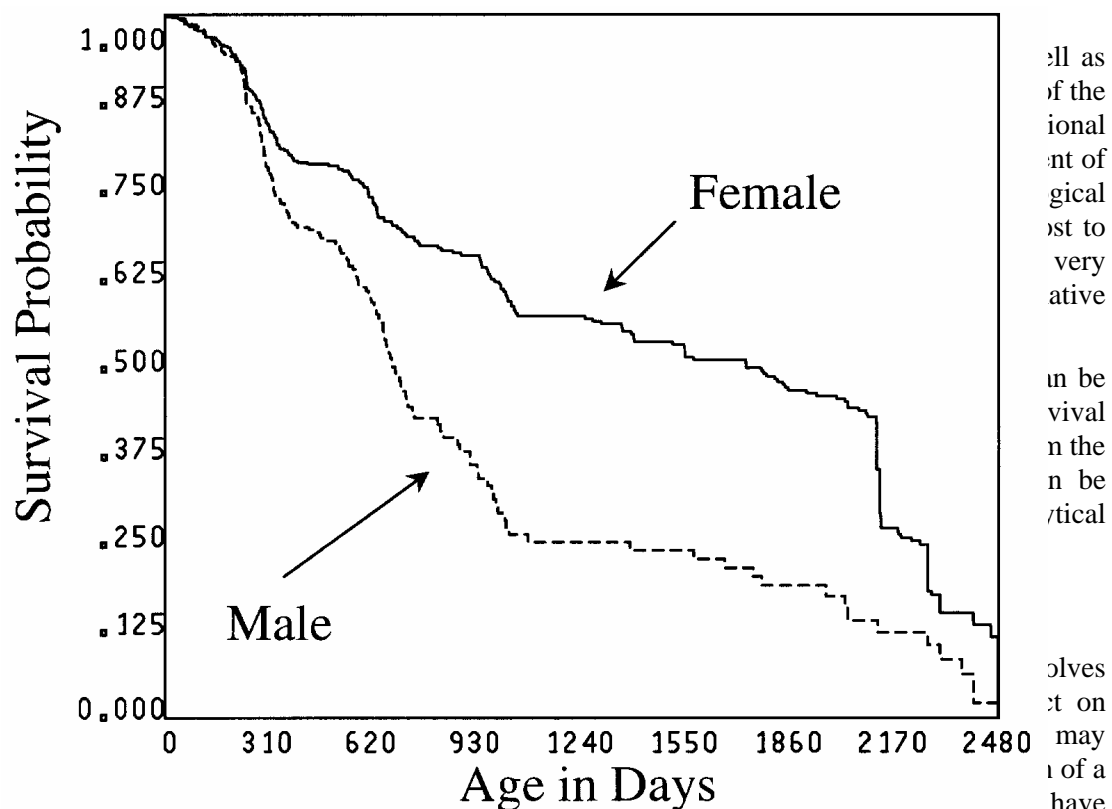
In a “classic” cohort study, a group of animals (cohort) which are exposed to a hypothesized risk factor are evaluated over time with a comparable group not exposed to the factor, to compare the rate at which the disease of interest develops in the two groups. In practice, there may be multiple cohorts, or cohorts may be extracted for analytical purposes from a population study, provided that this can be done without bias. Because in a cohort study animals are allocated to analysis groups on the basis of the putative risk factor and then subsequently observed through time for the occurrence of the disease, the strength of inferences which can be drawn from a cohort study is much stronger than from a case-control study, where animals are chosen on whether or not they have suffered the disease, and their background history is examined for apparently influential risk factors. Cohort studies have been used in the horse (Kobluk *et al*, 1990, 1991; Ross and Kaneene, 1996 a, b)

LONGITUDINAL POPULATION STUDY

This is a study of a population through time, in which individuals are identified and monitored through a sequence of observations. For example, blood samples may be taken monthly for serology or haematology, measurements of bone strength may be made every six months during the growth of foals, or reproductive performance and ovarian function of mares may be evaluated. The population may be sampled using stratified random sampling techniques to estimate population parameters, or the study may be of volunteer studs or training establishments. Studies have been conducted on racing populations (Peloso *et al* 1994), Physick-Sheard 1986 a, b; Physick-Sheard and Russell 1986) to evaluate performance, and studies have been made on wastage at various stages of the equine life cycle (Haas *et al* 1996; Jeffcott *et al* 1982; Rosedale 1989).

We are using a comprehensive longitudinal study method we call “health and performance profiling” to study diseases and disorders in about 1,000 horses under training, linking these to various measures of track performance with a view to determining factors which adversely affect racing performance, or conversely which are protective of ability. This study approach is resource-hungry, but provides comprehensive information on the same animals, and hence allows us to draw a range of conclusions about components of the causal web from within the one study. Moreover, a variety of specific studies can be nested within the overall design, providing a range of results from a single data collection effort. This is therefore a very useful technique, which is receiving increasing attention, although it should be conducted only occasionally, generating a range of hypotheses which can be tested in studies designed for the specific purpose.

Figure 3. Example of a survival analysis



been adequately characterised, an intervention study which changes the occurrence of a disease is the strongest single form of evidence that the factor which was varied is causally related to the occurrence of the disease or performance difference.

Analysis of such studies is typically by means of an appropriate form of the general linear model, such as analysis of variance or multiple linear regression. However a variety of other analytical approaches may also be used in appropriate cases.

DISEASE PROCESS STUDIES

In recent years we have applied epidemiological thinking to research on the pathogenesis of disease and investigation of the mechanisms underlying the relationship between disease and performance. While there is no standardised method for conducting such studies, the approach of investigating such issues by examining the dynamics of the process and progressively creating a “pathogenesis web” has proved quite rewarding in some diseases, and is illustrated in Figure 1.

MODELLING AND PREDICTION

We have worked extensively with the use of computer models to analyze diseases and evaluate the potential merits of control methods. Building such models is becoming steadily easier, from spreadsheets to compartment models to full simulation models, and they may have an application in studies of the horse. Predictive modelling can also be conducted using a purely empirical statistical approach rather than a mechanism-based approach, and various studies of this kind have been reported (Martin *et al* 1996, 1997; Martinelli *et al* 1996).

SYNTHESIS OF AN APPROACH

The discussion so far has outlined techniques, and provided illustrations of their application, including a summary of applications to the horse. We try always to apply these techniques to a particular problem in an organized fashion, to produce the required answers as efficiently as possible. When presented with a new problem, we will typically first characterise it through a case series study. Using the initial findings, we would conduct a case-control study to identify likely important risk factors which may be susceptible to control. We may then define the variables which require more robust evidence to support their influence, and evaluate this through a cohort study or intervention study. If we need to assess the dynamics of a disease in an animal population, we will commonly use a longitudinal study to do so, and if the issue justifies it we will use a cross-sectional study to define the importance of the condition across the full population of horses, or to assess the attitudes and understanding of the owners or managers by a questionnaire study. We can then bring our results together into a causal web or systems diagram such as Figure 1, and may delve more deeply into the pathogenesis of the condition if necessary.

Provided that we have our case definition well characterised and our investigational methods are sound, we do not need to know the causal agent or pathology of the disease in depth in order to provide answers on controlling the disease or improving animal performance - we are still able to provide epidemiological answers to the key questions.

REFERENCES

- Bailey CJ, Rose RJ, Reid SWJ and Hodgson DR, (1997). Wastage in the Australian Thoroughbred racing industry: a survey of Sydney trainers. *Aust vet J*, **75**: 64-66
- Bailey CJ, Reid SWJ, Hodgson DR, Suann CJ and Rose RJ, (1996). Risk factors associated with musculoskeletal injuries in Australian Thoroughbred racehorses, *Prev Vet Med*
- Dolvik, NI, (1994). Prevalence and risk factors in serous arthritis in the carpal joint of Norwegian cold-blooded trotters. *Vet Res Communications*, **18**: 281-288.
- Haas, SD, Bristol, F ad Card, C E, (1996). Risk factors associated with the incidence of foal mortality in an extensively managed mare herd.. *Can vet J*, **37**: 91-95.
- Jeffcott, L B, Rossdale, PD, Freestone, J, Frank, CJ and Towers-Clark, PF, (1982). An assessment of wastage in Thoroughbred racing from conception to 4 years of age.. *Equine vet J*, **14**: 185-198.
- Kaneene, JB, Miller, R, Ross, W and Gilbert, L, (1996). Patterns of health maintenance on Michigan equine operations. *Prev vet Med* **29**: 201-220.
- Kaneene, JB, Saffell, M, Fedewa, DJ, Gallagher, K and Chaddock, HM, (1997a). The Michigan equine monitoring system. I. Design, implementation and population estimates. *Prev vet Med* **29**: 263-275.
- Kaneene, JB, Ross, W A and Miller, R, (1997b). The Michigan equine monitoring system. II. Frequencies and impact of selected health problems. *Prev vet Med* **29**: 277-292.
- Kobluk CN, Robinson RA, Clanton CJ, Trent AM, Ames TR and Gordon BJ, (1991). Comparison of the exercise level and problem rate of 95 Thoroughbred horses: a cohort study. *Proceedings of the Annual Convention of the American Association of Equine Practitioners*, **36**: 471-475.
- Kobluk CN, Robinson RA, Gordon BJ, Clanton CJ, Trent AM and Ames TR., (1990) The effect of conformation and shoeing: a cohort study of 95 Thoroughbred racehorses. *Proc of the Annual Convention of the American Association of Equine Practitioners*, **35**: 259-274.
- Martin, GS, Strand, E and Kearney, MT., (1996) Use of statistical models to evaluate racing performance in Thoroughbreds. *J Amer vet med Assn* **209**: 1900-1906.
- Martin, GS, Strand, E and Kearney, MT., (1997) Validation of a regression model for standardizing lifetime racing performance of Thoroughbreds. *J Amer vet med Assn* **210**: 1641-1645.
- Martinelli, MJ, Freeman, DE, and Reid, SWJ., (1996) Analysis of performance parameters in the population of standardbred racehorses in the US (1984-1993). *Amer Assn Eq Practitioners* **42**: 174-5.
- Mohammed HO, Hill T, Lowe J., (1991) Risk factors associated with injuries in Thoroughbred horses. *Equine Vet J*, **23**: 445-448.
- Oikawa M, Ueda Y, Inada S, Tsuchikawa T, Kusano H, Takeda A., (1994) Effect of restructuring a racetrack on the occurrence of racing injuries in Thoroughbred horses. *J Eq Vet Sci*, **14(5)**: 262-268.
- Peloso JG, Mondy GD, Cohen ND., (1994) Prevalence of, and factors associated with, musculoskeletal racing injuries of Thoroughbreds. *JAVMA*, **204**: 620-262.
- Physick-Sheard, P W., (1986) Career profile of the Canadian Standardbred. I. Influence of age, gait and sex upon chances of racing. II Influence of age, gait and sex upon number of races, money won and race times. *Canadian Journal of Veterinary Research*, **50(4)**: 449-470.

Physick-Sheard PW and Russell M., (1986) Career profile of the Canadian Standardbred. 3. Influence of temporary absence from racing and season. *Canadian Journal of Veterinary Research*, **50(4)**: 471-478.

Reeves MJ, Gay JM, Hilbert BJ and Morris RS., (1989) Association of age, sex and breed factors in acute equine colic: a retrospective study of 320 cases admitted to a veterinary teaching hospital in the USA. *Preventive Veterinary Medicine*, **7(2)**: 149-160.

Ross, W A and Kaneene, JB., (1996a) An individual-animal-level prospective study of risk factors associated with the occurrence of lameness in the Michigan (USA) equine population. *Prev vet Med* **29**: 59-75.

Ross, W A and Kaneene, JB., (1996b) An operation-level prospective study of risk factors associated with the incidence density of lameness in Michigan (USA) equine operations. *Prev vet Med* **29**: 209-224.

Rossdale PD., (1989) The epidemiology of wastage in thoroughbred horses. *Pro Veterinario* **9(2)**: 5-6.

Wilson JH and Robinson RA., (1996) Risk factors for equine racing injuries. *Compendium of Continuing Education* **18**: 682-690.

MULTIVARIABLE AND MULTIFACTOR TECHNIQUES: AN INTRODUCTION

Stuart Reid

INTRODUCTION

In general, whilst there is a need to concentrate on getting the basics of data analysis correct, there will come a time when it will not be sufficient to deal with one factor or variable and its dependent outcome. Most diseases are multifactorial in nature and it stands to reason that we may wish to investigate the effect of one variable whilst controlling for the effect of all others. In order to provide a taste of the potential utility of such techniques we will discuss the steps involved in techniques from two different branches of epidemiology, observational and experimental. The techniques are:

1. multivariable logistic regression
2. analysis of variance (ANOVA) techniques

This course is too short to allow the methods to be dealt with in any depth but these notes together with the practical sessions should provide a useful overview. The section on logistic regression borrows heavily on Applied Logistic Regression by Hosmer and Lemeshow (Wiley, 1989) and Kleinbaum's Logistic Regression (1994). The ANOVA notes are based on Professor George Gettinby's lecture notes for Pharmacology students at the University of Strathclyde

MULTIVARIABLE LOGISTIC REGRESSION

The background

Mathematical and statistical modelling is an attempt to describe the relationship between a dependent or response variable (outcome) and a set of independent variables, also known as predictor or explanatory variables. In the context of biological relationships the outcome is very often continuous and may be related to a number of covariates. However, when the outcome is the occurrence of disease, the outcome is dichotomous (i.e. either disease occurred or it did not) but may still be related to a large number of possible explanatory variables. The fact that the dependent variable is binary lends the system to be described by a mathematical model which also has a dichotomous outcome and which provides a range of probabilities between the two extremes of one and zero. Such a model is found in the logistic function. Recently, the logistic model has become one of the most frequently used by quantitative biologists and epidemiologists, particularly where there are a number of explanatory variables involved, leading to a need to investigate the effect of each of these variables or factors individually whilst simultaneously controlling for all others in a multivariable analysis.

From an epidemiological point of view, any attempt to describe the relationship between two variables should be as simple as possible. First, the mathematical relationship should be easily understandable and interpretable, preferably by graphical means. Second, all known factors influencing the relationship should be accounted or excluded, the latter being the case if the effect of these variables is small and therefore may be considered insignificant. Wherever possible then, simplicity and inclusion of important elements should be the objectives of the modelling process at the same time as producing a model that fits the observed data.

The most basic of mathematical relationships is linear, of the form $y = \beta_0 + \beta_1 x_1 + \varepsilon$ in which the parameters of the slope, β_1 , and the intercept, β_0 , are readily interpretable. However, this form of relationship is rarely encountered in nature and transformations of biological data may be required to facilitate the implementation of a linear model in some way.

An important term in epidemiological studies is the cumulative incidence (rate) which can be regarded as the proportion of a fixed population that becomes diseased in a specified period of time. It can also be thought of as the probability of disease occurrence which, as a probability, has an upper limit of 1. In linear regression it is possible to have outcomes outside this upper limit and therefore we require a model in which the probability of disease is transformed to have a range from plus infinity to minus infinity. To overcome this problem two steps are required. First, the odds of the probability of disease is taken which then has a range from zero to infinity.

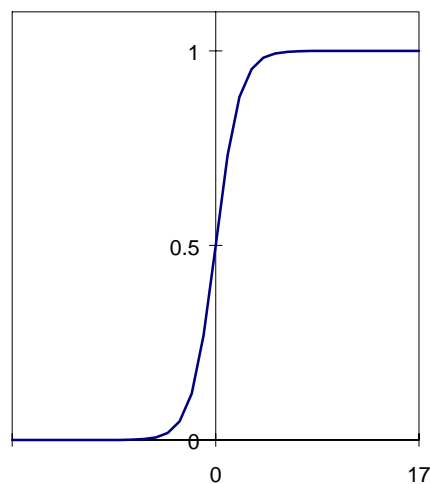
$$P(\text{disease}) / (1-P(\text{disease}))$$

Similarly, linear regression could yield negative values: this is overcome by taking the log of the odds:

$$\ln[P(\text{disease}) / (1-P(\text{disease}))]$$

What we have performed is a logit transformation and we now require a mathematical model that makes use of such a transformation. One such model is the logistic function. It is a mathematical model that has a range between 0 and 1 and is s-shaped which provides it with attractive properties for the epidemiologist. The function is defined as,

$$f(z) = \frac{1}{1 + e^{-z}}$$



The shape of the logistic function

Terms like lag, exponential growth and plateau phases may already be familiar to the statistician with an interest in the biological sciences but for the investigator modelling disease, risk and threshold of risk are important considerations. Not only does the logistic model conform to the mathematical constraints, the shape of the curve fits what epidemiologists consider the likely accumulation of risk; that there is a threshold before risk increases dramatically.

If the term z is then considered to be a linear combination of the independent variables such that:

$$z = \beta_0 + \beta_1 x_1 + \beta_n x_n$$

$$f(z) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_n x_n)}}$$

and if we consider that instead of a function of z , we refer to the last expression as a probability of the outcome X given the combination of all x s, we have an equation that starts to look useful. Note also that it does not matter what is on the right hand side of the equation, $f(z)$ will always be between 0 and 1.

In epidemiology we will be predominantly interested in the ratio of $P(X)$ to $P(1-X)$ or the relative risk which under certain circumstances is very similar to the odds ratio (i.e., when P is small).

Following on,

$$\begin{aligned} \frac{P(X)}{1 - P(X)} &= \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_n x_n)}} \div \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_n x_n)}} \\ &= e^{(\beta_0 + \beta_1 x_1 + \beta_n x_n)} \\ \ln \left[\frac{P(X)}{1 - P(X)} \right] &= \ln \left[e^{(\beta_0 + \beta_1 x_1 + \beta_n x_n)} \right] \\ &= \beta_0 + \beta_1 x_1 + \beta_n x_n \end{aligned}$$

Another way of looking at this and a way which will show us how useful the model can be is, to approach it from the classical two by two contingency table, and with a single dichotomous independent variable the individual cells would be:

	Factor present	Factor not present
Disease (cases)	$\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}}$	$\frac{e^{\beta_0}}{1 + e^{\beta_0}}$
Not diseased (non cases)	$\frac{1}{1 + e^{\beta_0 + \beta_1}}$	$\frac{1}{1 + e^{\beta_0}}$

The odds ratio for the factor above would be:

$$\begin{aligned} \text{odds ratio} &= \frac{\left[\frac{e^{\beta_0 + \beta_1}}{1 + e^{\beta_0 + \beta_1}} \right] \left[\frac{1}{1 + e^{\beta_0}} \right]}{\left[\frac{e^{\beta_0}}{1 + e^{\beta_0}} \right] \left[\frac{1}{1 + e^{\beta_0 + \beta_1}} \right]} \\ &= \left[\frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} \right] \\ &= e^{\beta_1} \\ \log \text{ odds ratio} &= \ln(e^{\beta_1}) \\ &= \beta_1 \end{aligned}$$

So, it can be seen that the exponent of the coefficient for a dichotomous variable is equivalent to the log odds. Also note that the log of the odds ratio of $P/(1-P)$ is equivalent to the log of P minus the log of $(1-P)$. In other words the coefficient, β_1 , is the log odds ratio and the difference in probability on a log scale or the logit difference.

As regards the incorporation of additional factors, as discussed earlier, the logistic model does not constrain us in the number of variables that may be included. The important thing is that we can include any number of independent variables and the relationship holds true....exponentiate any coefficient and the odd ratio is computed. Parsimony however requires attention to be paid to significance within the model and is dependent upon model building strategy.

The essence of logistic modelling

With an understanding of why the model should be useful to us we can now quite quickly see how we would go about performing a logistic regression analysis.

1. Classify the animals according to disease status.
2. Reduce the number of variables and factors that are to be included in the model. This can be done by screening the variable using an appropriate test; 2 sample t tests for continuous variables, chi-square tests for categorical. Discard any variable that is not significantly associated with disease at $P < 0.25$.
3. Include all remaining variable in a multiple logistic regression model. Run the model.
4. Refine the model by sequentially removing the variables and looking for an improvement in the model by inspection of the deviance (this will be dealt with in the practical session).
5. Inspect the model for absolute goodness-of-fit.
6. Perform model diagnostics (beyond the scope of this course).

The final model will then allow us to investigate the influence of each of the variables on the outcome whilst controlling for the others. The technique is now commonplace but whilst it may be easy to perform, there are dangers and it should be used with caution.

ANALYSIS OF VARIANCE

Introduction

Computationally somewhat easier to perform, ANOVA techniques are multifactor technique that are used to analyse data from experimental epidemiological investigations such as clinical trials. As above I assume a level of knowledge regarding simple techniques for data summarising and simple statistical tests for comparing different groups, like Students t-tests and chi-square tests. This typically is the first step we undertake when performing population epidemiological or pharmaco-vigilance studies. However, we can be more proactive in designing studies in order to maximise the utility of the findings. For the analysis of continuous variables this often leads to the use of particular *experimental designs*. Let us look at some human examples.

One-factor design

Suppose we wish to investigate the effect of the three drug treatments X,Y and Z on blood pressure in subjects suffering from clinical hypertension. A number of subjects are recruited to the study and *randomly* assigned to receive treatment X or Y or Z. In order to make the comparisons equitable we design the study so that each treatment group has the same number of subjects. If there are 12 subjects available, each group will have 4 subjects and the data will be of the form:

Treatment Group		
X	Y	Z
211	194	180
215	180	195
194	180	173
205	187	172

Entering these to our statistical package (or spreadsheet) and depending on whether the corresponding datum entered in C1 belongs to treatment group X or Y or Z we have:

	C1 bloodpr	C2 GROUP
1	211	1
2	215	1
3	194	1
4	205	1
5	194	2
6	180	2
7	180	2
8	187	2
9	180	3
10	195	3
11	173	3
12	172	3

This situation is very similar to two-sample comparisons only now we have three samples to compare and we wish to know “do any of them come from populations with different means?”. Such an experimental design can have any number of treatment groups and is analysed using analysis of variance (ANOVA) and the F-test which is a variation on the t-test.

ANOVA compares the variability between the treatment groups with the variability within the treatment groups. If the ratio of the variability between the treatment groups to the variability within the treatment groups is improbably large i.e. typically only occurs with a p-value of less than 0.05 this provides evidence that treatment means differ more than the typical variation from one subject to another within a treatment group, and we say that there is a significant difference. If the F value is small and close to 1 (and the associated p-value large) we need not proceed any further as the conclusion is that there is no evidence of a significant difference.

The output in Minitab for our example is as follows:

MTB > oneway c1 c2

One-Way Analysis of Variance

Analysis of Variance on bloodpr

Source	DF	SS	MS	F	p
group	2	1543.5	771.7	9.60	0.006
Error	9	723.5	80.4		
Total	11	2267.0			

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev	
1	4	206.25	9.14	-----+-----+-----+----- (-----*-----)	
2	4	185.25	6.70	(-----*-----)	
3	4	180.00	10.61	(-----*-----)	

Pooled StDev = 8.97

180 195 210

MTB >

Here we see that the F-value has a p-value of 0.006 and so there is clear evidence of a significant difference.

Multiple Range Test

When there is evidence of a significant difference between the groups then we must proceed to find out which groups are different to each other. This is carried out using a Multiple Range Test which is essentially the use of pairwise two-sample t-tests to test if group X is different to Y, group X is different to Z and group Y is different to Z. There are several such tests available and in our example choosing Fisher's pairwise comparison test it lists the 95% Confidence Intervals for the difference between the means of all pairs of treatments groups:

MTB > oneway c1 c2;
SUBC> Fisher 5.

One-Way Analysis of Variance

Analysis of Variance on bloodpr

Source	DF	SS	MS	F	p
group	2	1543.5	771.7	9.60	0.006
Error	9	723.5	80.4		
Total	11	2267.0			

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev	
1	4	206.25	9.14	-----+-----+-----+----- (-----*-----)	
2	4	185.25	6.70	(-----*-----)	
3	4	180.00	10.61	(-----*-----)	

Pooled StDev = 8.97

180 195 210

Fisher's pairwise comparisons

Family error rate = 0.113
Individual error rate = 0.0500

Critical value = 2.262

Intervals for (column level mean) - (row level mean)

	1	2
2	6.66 35.34	
3	11.91 40.59	-9.09 19.59

MTB >

Those confidence intervals not containing 0 indicate that the treatment groups are different as 0 cannot be ruled out as a possible value for the difference between the treatment means

which would indicate the means to be the same. From above we see that there is evidence that treatment groups X and Y are different (confidence interval 6.66 to 35.34) and treatment groups X and Z are different (confidence interval 11.91 to 40.59), whereas treatment groups Y and Z are not different (confidence interval -9.09 to 19.59). From inspection of the means it is apparent that blood pressure levels are significantly lower for patients on treatments Y (mean 185.25) and Z (mean 180.0) compared to treatment X (mean 206.25).

Two factor design

Suppose we suspect the sex of the patient may have a bearing on the blood pressure levels when treatments are compared and this would be important in treating subjects. We can *design* the study to have equal numbers of male and female subjects in each of the treatment groups:

		Treatment Group		
		X	Y	Z
Sex	Male	211	194	180
		215	180	195
	Female	194	180	173
		205	187	172

We are still able to compare the treatment groups as before by using the ANOVA and the F-test. Any sex effect will be balanced out as there is an equal number of males and females in each of the groups. Moreover, we can also use the ANOVA in a similar fashion to test for any difference between Male and Female subjects. This is a two-factor design and the ANOVA provides an F-test for treatment group differences and an F-test for sex group differences. To implement the analysis of the design we must enter a new data column containing the coding for each blood pressure value according to the designated sex i.e. 1 for male and 2 for female:

	C1 bloodpr	C2 GROUP	C3 SEX
1	211	1	1
2	215	1	1
3	194	1	2
4	205	1	2
5	194	2	1
6	180	2	1
7	180	2	2
8	187	2	2
9	180	3	1
10	195	3	1
11	173	3	2
12	172	3	2

Now we are able to carry out the statistical analysis using the following commands to get:

```
MTB > anova c1 = c2 c3;
SUBC> means c2 c3.
```

Analysis of Variance (Balanced Designs)

Factor	Type	Levels	Values
group	fixed	3	1 2 3
sex	fixed	2	1 2

Analysis of Variance for bloodpr

Source	DF	SS	MS	F	P
group	2	1543.50	771.75	16.16	0.002
sex	1	341.33	341.33	7.15	0.028
Error	8	382.17	47.77		
Total	11	2267.00			

MEANS

group	N	bloodpr
1	4	206.25
2	4	185.25
3	4	180.00

sex	N	bloodpr
1	6	195.83
2	6	185.17

MTB >

This identifies the F-value for GROUP to be significant as before ($p=0.002$) and also the F-value for SEX to be significant ($p = 0.028$). Consequently there is evidence that not only are there treatment differences but males respond differently to females. On average males have an average blood pressure level of 195.83 whereas females have an average level of 185.17 and the difference is statistically significant. This may have implications for using the treatments in the population at large.

We therefore see that by using the same number of experimental subjects as in the one-factor design we can answer both questions about differences between treatment groups and also differences between sexes by designing the study as a two-factor design. By using an appropriately planned experimental design we have gained extra knowledge at no extra cost!

Other experimental designs

The one and two-factor designs can be extended to three-factor, four-factor designs and so on. For example, we could introduce an additional factor diet and assign each of the two subjects in each of the above treatment-sex combinations to receive either diet 1 or diet 2. This would become a three-factor design and the analysis of variance would provide three F-tests, one for differences amongst treatments, one for differences between sexes and one for differences between diets.

In our example the GROUP factor had 3 levels as there were 3 treatments. The SEX factor had 2 levels as there were males and females. In practice, each factor can have any number of levels.

Experimental designs can be used to test for different response patterns across groups known as *Interactions*. Designs are also widely used to test for bioequivalence of compounds and effects over time using Latin Square and Repeated Measures Designs. However, for the methods to be appropriate it is generally required for the data to come from normal or almost normal distributions and so you cannot use it on all types of measurements.

The two simple designs we have discussed should help you with your project work. Remember, two key features of the designs are the principles of *randomisation* and designing your study to produce *balanced* data.

MAKING THE MOST OF LARGE CLINICAL DATASETS

Kathryn Knox

INTRODUCTION

As a result of the technological revolution, computerised recording and retrieval of large clinical datasets suitable for quantitative analysis is now possible. Although the ideal electronic medical record is yet to be developed, several authors have suggested that establishing standard computerised record-keeping systems in general practices would have significant implications for obtaining information about the practice population. With increasing use of hospital databases in veterinary research, the requirement for assured reliability of electronic data, appropriate application of statistical techniques, and validity of study conclusions are of utmost importance. In dealing with large clinical datasets, the ultimate aim is to summarise and present the data in a more useful format.

VETERINARY CLINICAL DATA

In the field of medicine, patient-based data storage at its most basic level is in the form of a “medical record”. In both human and veterinary medicines, the importance of a well-maintained medical record is paramount. The veterinary medical record is maintained for a number of reasons. The primary functions of a medical record include providing documentation of a patient’s health status and care; providing a means of communication among veterinarians; and, teaching and research purposes. On an individual level, medical records may be used for teaching students about a particular disease condition or the approach to an unwell animal by following through a medical record and tracing the course of the events. Collectively, medical records can be used for epidemiological studies.

The veterinary medical record may be as simple or as complicated as is necessary. Currently, in the United Kingdom there are no official detailed guidelines as to how an individual medical record should be structured. Thus, the details of the record depend largely on the requirements and policies of individual practices or hospitals. For the maintenance of useful records, however, the types of information which may be stored are numerous. The animal identification itself should contain the following information: name, species, breed, date of birth or age and gender. Basic details to identify the animal’s owner may also be required, both for contact purposes and also to give an indication of the geographical location of the animal. Other types of information which may be maintained within a medical record relate more specifically to the health status of the animal at any particular time. Most of the domestic species, both companion animal and production animal, undergo veterinary inspection as healthy animals, whether for neutering, vaccination or health screening. Thus, even those animals which remain “healthy” throughout their lives have a medical record. It is during periods of illness that the medical record becomes potentially complicated. The types of information which may be collected include: history, clinical examination, details of further tests, test results, differential diagnoses, treatment instigated, prognosis and outcome. These details may be obtained during one visit, or, more usually, over a number of visits. From this very brief outline of the types of veterinary clinical data which may be collected, it is apparent that, over time, large clinical datasets may be established in both veterinary practice and hospital environments.

APPROACH TO LARGE CLINICAL DATASETS

Which data?

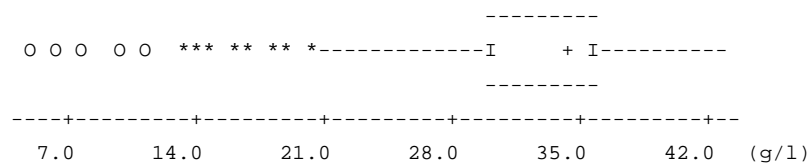
Usually, large clinical datasets are divided into smaller subsets for investigation. Although the most likely reason for this may be that the investigator has a specific interest, for example in the diagnosis of hepatopathy in horses, any attempts to use *all* the information from a large clinical dataset are likely to be futile. The focus of any investigation into a large clinical

dataset, therefore, is dependent on the area of interest of the investigator, as well as on the types of data available. However, certain aspects of clinical dataset investigation are likely to be common to most circumstances.

Veracity of data

The reliability of the electronic data is of utmost importance. For continuous data, for example biochemistry results or haematology results, statistical outliers should be identified by obtaining boxplots for the parameters, as illustrated in Figure 1.

Figure 1 Example of a boxplot, displaying the distribution of albumin results (g/l) for horses which presented to the University of Glasgow Veterinary School hospital.



A boxplot provides a pictorial representation of the distribution of the results, and helps identify extreme values. In the statistical software package Minitab which was used to generate Figure 1, * represents a possible outlier, while 0 represents a probable outlier. All values identified as outliers should be checked for biological possibility and also by referral to hard copy. Similarly, text entries, such as presenting signs or diagnoses, must be verified. One of the most common problems associated with textual entries in large clinical datasets is the lack of standardisation. For example, one clinician may enter “Cushing’s disease”, while another enters “Cushingoid”, and another still may enter “hyperadrenocorticism”. Similarly, “liver” may be entered by some clinicians, while “hepatic” is entered by others. Some institutes have addressed this problem by adopting a standard nomenclature, such as SNVDO, Standard Nomenclature of Veterinary Diseases and Operations, where all diagnoses are coded according to a specific numerical format. However, if such standard coding has not been adopted, diagnosis data must be manually standardised by the investigator, possibly with the development of an appropriate numerical coding system.

Summarising data

When the validity of all data for investigation has been ensured, appropriate statistical techniques may be employed to summarise, analyse and present the data. Application of the statistical techniques is dependent on the data type under investigation and on the “questions” being asked of the data. Statistical techniques may range from simple descriptive statistics such as means, standard deviations, medians, quartiles or frequencies, to more complex statistical procedures such as cluster analysis or multivariable logistic regression. Discussion of the details pertaining to different statistical techniques which may be applied to clinical datasets is beyond the scope of this manuscript, however, most of the simple data presentation and summary methods, such as graphical presentations of data, i.e. boxplots, bar charts, histograms, and basic descriptive statistics may be readily undertaken in proprietary spreadsheet or statistical software packages.

Software Packages

When dealing with large clinical datasets, the ability to manipulate data is vital. The use of a spreadsheet package, such as Excel, facilitates data handling and ultimately allows appropriate presentation of the data. The ease with which data may be transferred from a database package to Excel depends on the original database software, but the “Text Import Wizard” within versions 5.0 and above of Excel, automatically detects if a selected file is not in Excel format, and the user will be prompted to provide information with respect to data

type, for example, detailing whether the data are delimited, and, if so, by what character, e.g. commas or tabs; thereafter, the data are automatically imported to an Excel spreadsheet within appropriate columns. It is then advisable that the file is subsequently saved in Excel format. The file may then be manipulated readily using all spreadsheet functions. If more complex statistical analysis is to be undertaken on the data, such as hypothesis testing or regression analysis, the data may be exported to an appropriate statistical software package, such as Minitab or Statistix.

Drawing Conclusions

Finally, with respect to drawing conclusions from the investigation of any large clinical dataset, it is important to be aware of issues which may affect the interpretation of the results. For example, investigation of data from a hospital based population will yield results which may not be applicable to a practice population. However, as long as due care is observed, investigation of large clinical datasets may generate important results and add to medical knowledge as a whole.

An example of how these techniques may be useful is illustrated using the large computerised database maintained at the University of Glasgow Veterinary School hospital. The investigation of this particular database is outlined, highlighting statistical approaches adopted, particular study considerations and implications of the findings.

INVESTIGATION OF GUVS HOSPITAL DATABASE

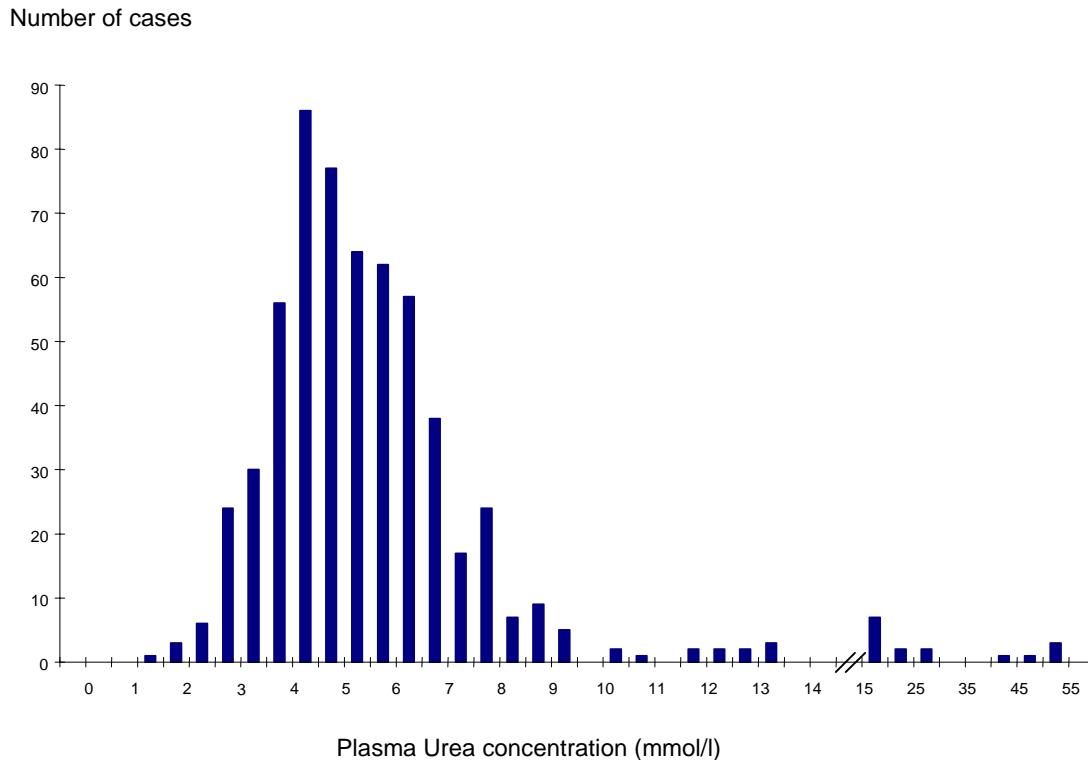
The University of Glasgow Veterinary School (GUVS) functions as a veterinary referral unit, as well as a teaching and research institute. It is responsible for the referral care of a wide variety of species, but most usually dogs, cats, horses, cattle and sheep. For most of the species, clinical cases are referred from general veterinary practices in the surrounding geographical area. Each animal which presents to the hospital undergoes comprehensive historical and clinical examination, further ancillary clinical testing, if required, and ultimately therapy and follow-up investigations. GUVS has maintained a large computerised data recording system for a number of years. The system, containing information on over 25,000 animals, is comprised of a number of databases containing signalment, clinical, biochemical, haematological and pathological information. This large clinical dataset offers a rare and potentially valuable resource.

The computerised database was interrogated to identify those horses which had plasma biochemistry tests carried out as part of their work-up. The biochemistry results for these cases were extracted from the database and imported to an Excel spreadsheet for subsequent validation and analysis. For each horse with at least one panel of biochemistry results, the clinical diagnosis associated with that hospital visit was also retrieved from the database.

The dataset was summarised using descriptive statistics and graphical presentations, such as that shown in Figure 2. By applying a combination of percentile analysis and conditional probability techniques to the biochemistry and diagnosis data, a means for the objective interpretation of clinical biochemistry data was developed. From knowing a parameter concentration, a clinician was able to establish whether a value was abnormal, the *degree* of abnormality and the most likely associated diagnoses. Furthermore, a “Biochemical Factor” indicated how many times more likely a particular disease was, given a particular biochemistry parameter concentration. Utilising the data from within the large dataset therefore yielded valuable information.

However, such information is only truly of value if it made is readily available to interested parties. The results from the interrogation of the GUVS hospital database were therefore used to form the basis of a biochemistry decision support system.

Figure 2 Distribution of plasma urea results (mmol/l) from equine referral cases seen at GUVS over the period of the study.



Issues Pertaining To The Hospital Dataset Investigation

The investigation of any large clinical dataset must be undertaken with care. Before conclusions were drawn from the results of the interrogation of hospital biochemistry data, a number of issues had to be considered. Many unwell horses which were investigated at GUVS required more than one biochemistry test. This may have been for several reasons including monitoring a case over time or revisitation to the hospital for an unrelated reason. Only the first biochemistry sample result for each case was included in the study in an attempt to remove this source of potential bias. A further potential complication was the variation between results when different biochemistry methodologies were employed. It is recognised that when different techniques are used to perform biochemistry assays, results can vary. During the study period, two different biochemistry analysers were used. For this reason, the data were divided into two appropriate groups, and then standardised by percentile analysis.

The diagnoses used in the study were often tentative clinical diagnoses. This was a very important consideration, because it highlighted the possibility of a circular argument. If a diagnosis was made with the aid of a panel of clinical biochemistry results, then the biochemistry results may in turn reflect that diagnosis. It could therefore be argued that some of the results of the GUVS investigation were, in fact, based on opinion, as opposed to being completely data derived. Although an important recognition, this circular argument was unlikely to have influenced the results to a large extent because most of the diagnoses will have been made using all the information available to the clinician, rather than relying on any one parameter result. This leads to another important consideration, namely that biochemistry parameters were examined on an individual basis. The next stage of the study could involve combining biochemistry parameters and performing multivariate and multivariable analyses. However, given that many statistical approaches assume independent variables, appropriate statistical methods must be chosen to ensure the valid continuation of the study.

Finally, it must be appreciated that the population under study was an unwell referral population and it must be regarded as a biased population. Many cases which present to general veterinary practice, for example straight forward spasmodic colic, may never need referral to a veterinary hospital, and thus would be under-represented in a hospital population. In contrast, Cushing's disease may be over-represented. Therefore, the results from this study, although valuable, may not be applied directly to the general practice situation. However, the statistical techniques described in this study could be similarly applied to a general practice population, with potentially important results.

Implications Of The Findings

In veterinary medicine, ancillary clinical testing has developed enormously in recent years. In particular, clinical biochemistry test results are now available to clinicians through the development of analysers designed for in-practice use. Whilst this is certainly beneficial, such test results are only truly of value when correctly interpreted. However, interpretation of such clinical biochemistry results remains rather subjective. Evaluation is essentially based on the experience of the individual clinician, who may refer to means and ranges on the premise that the data are normally distributed, an assumption which may not always be appropriate. Adoption of percentile analysis and probability techniques as discussed above allows the representation of clinical biochemistry values to be simplified at several levels. A parameter concentration may be described simply as falling within the top one or five per cent of all values achieved for horses at GUVS, with the absolute value becoming less necessary. Expression of biochemistry results in percentile form also redresses any possible confusion over units and, importantly, it allows comparison of biochemistry results which have been obtained using different biochemistry analysers. In this study, diagnoses were correlated with the biochemistry results. A clinician thus could appreciate whether a value was abnormal, the degree of abnormality, and also the most likely associated diagnoses. Furthermore, application of probability methods allowed clinicians to determine how many times more likely a particular diagnosis was, than before any information had been available.

A genuine understanding of disease processes and pathophysiological disturbances best equips a clinician to interpret clinical biochemistry data. However, the basic veterinary course is continually expanding, and time allocated to the teaching of clinical biochemistry testing and interpretation may, as a consequence, be diminished. The necessity for improving the approach to the interpretation of clinical biochemistry is thus of particular importance for the future. Appropriate application of the percentile and probability techniques applied to an appropriately defined population may offer a means for such interpretative support.

Conclusion

In the example involving the GUVS database, plasma biochemistry results from a putatively unwell referral equine population were investigated using percentile analysis and conditional probability methods to quantify associations between individual biochemistry parameter concentrations and clinical diagnoses. The results of the analysis of the dataset were used to form the basis of a biochemistry decision support system. Through a user-friendly computer interface, clinicians are able to objectively interpret clinical biochemistry data. The system is based on clinical data and provides an example of making the most of a large clinical dataset. In the early stages of this study, lots of individual "pieces of information" were available, that is, thousands of numbers representing biochemistry parameter concentrations and hundreds of clinical diagnoses pertaining to horses. On their own, these data provided only limited information. However, once collated, simply summarised and represented in a more visually appealing format, the information value of these data was realised.

Although the investigation of a specific clinical dataset has been described, with careful consideration, similar approaches may be applied to other such datasets. The investigation of any large clinical dataset must be undertaken with appropriate attention to particular

circumstances pertaining to the individual situation, and this should ensure not only that the dataset may be used to its full potential, but also that any possible limitations be appreciated.

SENSITIVITY, SPECIFICITY AND PREDICTIVE VALUES

Stuart W.J. Reid

INTRODUCTION

Whatever branch of equine research one may be involved it is inevitable that there will be a requirement at some stage to perform some kind of test which classifies animals into negative or positive, high or low, severe or moderate. Abramson (1994), in his excellent book "Making sense of data" quotes the Duchess from Lewis Carroll's "Alice's adventures in Wonderland" saying "Never imagine yourself not to be otherwise than what it might appear to others that what you were or might have been was not otherwise than what you had been would have appeared to them otherwise." The confusion that surrounds the interpretation of clinical tests is perhaps even more profound than Alice's dilemma and in most instances clinicians are required to make judgments based on faith rather than facts.

When one is teaching the basics of epidemiological methods to veterinary and medical students and specifically, the meaning and utility of measures such as sensitivity, specificity and predictive values the emphasis is generally placed on the results generated by the use of ancillary investigations such as clinical biochemistry and haematology. However, an appreciation of the application of these terms to medicine in general perhaps allows a better grasp of the whole diagnostic work up rather than solely the performance of laboratory based tests.

We have recently moved to re-introduce the terminology to our clinical students at the time they are standing next to the animal, stethoscope or thermometer in hand, anxious to arrive at a definitive diagnosis as quickly as possible. Many things in life are uncertain and it is critically important that the student realises that the hands-on diagnostic techniques he or she may be using are far from perfect, all the more so as they are generally being applied by one with fledgling skills and little experience upon which to base comparative judgements. In short, some of the time clinical measurements will be inaccurate, lacking in precision or just plain wrong. The same remains true for those of us involved in clinical research.

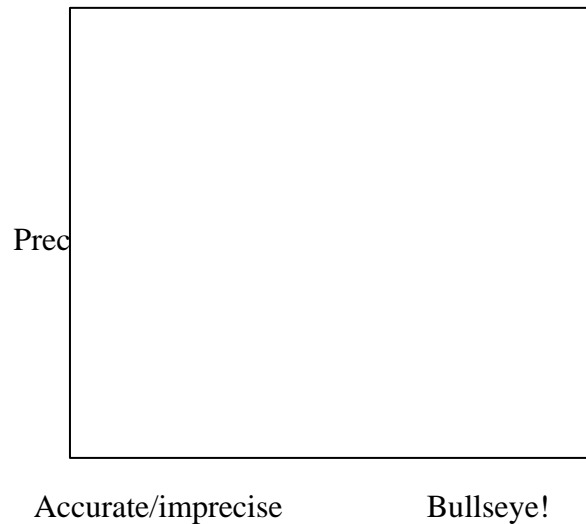
An appreciation of the fallibility and variation in the ability to hear murmurs, identify neurological deficits or fail to shake down a thermometer then facilitates the concept that laboratory results, however complex or expensive, may not always reflect what we shall call "the true state of nature". For instance, take the example of a mare diagnosed as "not pregnant" that 10 months subsequently produces a foal: It matters little whether it was a rectal examination or a PMSG assay that produced the result, the true state of nature was that the mare was pregnant.

The second important issue that must be grasped is that tests are used day in, day out, on many animals, such is the nature of veterinary clinical practice, but what is of *real* interest is how the test performs on balance. Every clinician, researcher or assayer would be forgiven the occasional mistake but were that mistake to happen in every pregnant mare examined then a lawyer would probably be the most useful ally.

THE PERFECT TEST

What then are the characteristics of the perfect test? Consider the analogy of the dart board, adapted from Kerr (1989).

The best scenario is obviously to have accuracy and precision but the single most important quality is that related to precision. If a test is not consistent, it is of very little use to us as we will never be comfortable in the knowledge that each test result must be interpreted differently. Inaccuracy that accompanies precision is less of a problem because we can apply a correction factor that will apply in all cases. In the case of a laboratory test we can think of these properties as being the responsibility of the lab, related as they are to the generation of the individual results from occasion to occasion.



SENSITIVITY AND SPECIFICITY

We must now think of how the test performs in a population of animals. Consider the 2 by 2 table:

	True state of nature		Total
	Positive	Negative	
Test +ve	a	b	a + b
Test -ve	c	d	c + d
Total	a + c	b + d	a + b + c + d

The contents of the cells “a”, “b”, “c” and “d” can be thought of as measures of the test’s performance in relation to the true state of nature. They are “true positives”, “false positives”, “false negatives” and “true negatives”, respectively, and a more intuitive way of looking at the table above is:

	True state of nature		Total
	Disease present	Disease absent	
Test +ve	TP	FP	TP + FP
Test -ve	FN	TN	FN + TN
Total	TP + FN	FP + TN	TP+FP+FN+TN

where TP is a true positive and FN is a false negative and so on.

So now we have a means of classifying the results of our tests in relation to a gold standard of some kind; this may be the definitive diagnosis, post mortem findings or the results provided by a proven quality test already in use. There are a number of measures we can derive from the tables above now that we can see their biological interpretation. The first of these is sensitivity. Sensitivity is an assessment of how the test performs in the diseased population:

how many of the diseased animals are we able to identify as diseased using our test?

$$\text{Sensitivity} = a / (a + c) = \text{TP} / (\text{TP} + \text{FN})$$

In contrast, specificity is an assessment of how well the test performs in the healthy part of the population: how many of the healthy animals would our test correctly identify as healthy?

$$\text{Specificity} = d / (b + d) = \text{TN} / (\text{TN} + \text{FP})$$

We can also provide an overall measure of the accuracy of the test (some call it efficiency) and this is “what proportion of the animals are correctly classified as being diseased or healthy, using our test?”

$$\begin{aligned} \text{Accuracy} &= (a + d) / (a + b + c + d) \\ &= (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \end{aligned}$$

OTHER ISSUES

However, to some extent we have been looking at things the wrong way around. We have asked the question “what is the probability that the animal’s disease status reflects the test result”. This is an important consideration but more important for the clinician is, “what is the probability that the test result reflects the disease status of the animal.” These may sound the same but a nice example from criminal law demonstrates the difference. During the much publicised trial of O.J. Simpson two years ago, one of the lawyers for the defence, Dershowitz, implied that “1/10 of 1 percent of wife batterers go on to murder their wives”. This may well be the case but in the journal *Nature*, Good wrote the correct reply to a flawed assumption. He stated that in cases where there was a murdered battered wife, the battering husband was found to be guilty in 1/2 cases.

So what is the difference? One is the probability of the woman being murdered by her husband given that she has been battered by him; the other is given that wife has been battered and is murdered, the chances are 1 in 2 that the husband committed the crime.

PREDICTIVE VALUES

In a similar fashion, the assessment we are really interested in is what is the probability that the test result reflects the disease status of the animal. This depends upon the prevalence of the disease as we shall see.

Suppose that plasma total protein concentration was being assessed as a test for medical colic.

	True state of nature		Total
	Medical colic	Not medical colic	
High protein	15	1	16
Low protein	5	10	15
Total	20	11	31

We can readily calculate that, for these data, the sensitivity of the test is 15/20, 75%, and the specificity is 10/11, 91 %. It can also be noted that the prevalence of medical colic is 20/31, 65%. However, in a similar study in a much larger equine practice the prevalence is much lower, at 8%.

	True state of nature		Total
	Medical colic	Not medical colic	
High protein	15	20	35
Low protein	5	210	215
Total	20	230	250

It is evident that the sensitivity and specificity of the test is the same as in the first population, but the importance of a positive test in the first case is arguably more informative than a positive result in the second case. The term we use for this the positive predictive value of the test and it is related to the prevalence of disease. Predictive values are calculated for both positive and negative results:

Positive Predictive value = $a / (a + b) = TP / (TP + FP)$

Negative Predictive value = $d / (c + d) = TN / (TN + FN)$

Practice 1: Positive Predictive value = 94%
 Negative Predictive value = 67%

Practice 2: Positive Predictive value = 43%
 Negative Predictive value = 98%

The reason that we have looked at these examples is to try and emphasize the need for each test to be assessed in the circumstances in which it will be used. False negatives and false positives are important at different times: Screening animals for presence of disease, screening animals for absence of disease or using tests on individual cases all require different approaches and it is critical that if the results are to be interpreted appropriately, the correct assumptions are adopted.

EXAMPLES

Murphy et al. (1997) assess the use of a modified oral glucose tolerance test as an indicator of small intestinal pathology in horses. The unmodified test is in common use at referral centres but the administration of a glucose solution *per os* and the subsequent blood sampling at approximately 20 minute intervals for the ensuing six hours precludes its use in general practice. Murphy demonstrated that by sampling only at 120 minutes and setting the test pass criteria at an increase in blood glucose of 20 % above basal glucose concentration, the sensitivity and specificity of the test were 90% and 64%, respectively, figures comparable with the unmodified test.

So much for sensitivity and specificity measures that are reasonable to use when they are applied to similar populations to that used for the test assessment. Cohen and Mackay (1997) elegantly demonstrated the problems associated with immunoblot testing of CSF fluid for Equine Protozoal Myeloencephalitis (EPM). The issues of relevance are that the test is known to generate erroneous results (for a number of potential reasons) and also that some diseases may be confused with EPM. The combined effect of these factors leads to

1. The possibility of false positives
2. Incorrect assumptions about the prevalence of the disease

What does this mean? Essentially in areas of low true prevalence, despite the fact that sensitivity and specificity of the disease may be good, the positive predictive value of the test is very poor.

SERIAL VERSUS PARALLEL TESTING

Usually when confronted by a clinical case that requires some form of testing to be undertaken in order to make a diagnosis, a number of tests will be run at the one time. This *parallel* testing will include both clinical and laboratory based tests. The latter in particular will be composed of batch or profile type testing. What is happening is a rapid assessment of the case against a perceived standard of normality or health and as such we are trying to rule out disease; the negative test results have the greatest predictive value.

Serial testing is one test applied after another in a deductive work up. This, contrary to parallel testing, increases specificity and is used when we are trying to rule in a disease. The most informative results are the positive ones but there is a chance that disease will be missed due to decreased sensitivity.

A third “screening” approach is retesting negative test result animals. Here again disease is being ruled out and sensitivity is maximised.

TRUST

As can be seen we must be aware of the strengths of the different strategies when we commence testing and the strategy will be different from situation to situation. Underlying all of these approaches is the basic assumption that what we receive by way of results is reliable and, as we have seen in the instance of lab based tests, the allocation of positive or negative is achieved by a means in which we have confidence.

When lab tests such as clinical biochemistry and haematology are being used the pivotal issue is “what is abnormal?”. In general, we take our guidance from the reference range, that is the mean value plus and minus two standard deviations, such that the normal limits cover 95% of the distribution of values from healthy animals. It therefore stands to reason that there is 5% chance that a healthy animal will be classified as diseased or abnormal on this basis. The more tests we do the greater the chance that at least one of those tests will be abnormal. The reference range is thus critical to our interpretation of these lab based data. But more important than this is the realisation that tests are never applied in isolation and should always be assessed in the light of other available information

So finally, what are the questions one should be asking of a laboratory? All of the measures we have talked about are important; sensitivity, specificity, and predictive values. However, probably the most important question for any researcher to ask is “what is the reference range for the test in *that* laboratory”. If the lab is unable or unwilling to provide this information it is likely that there has been no reference range established in the laboratory for that particular assay or test. This may mean that the lab expects the clinician to interpret the result by reference to the “normal” ranges for that analyte in that species, found published in text books. Second it may be the case that the lab is quoting a reference range that has been generated in a different laboratory albeit for that test protocol. However, hopefully it is now clear that neither of these situations allows the clinician to establish exactly where the result he or she has in hand relates to all the measures we have mentioned. Sensitivity, specificity and predictive values mean nothing if one does not have faith in the lab’s classifying normal as normal and abnormal as abnormal.

SOME FINAL COMMENTS

Although the examples that have been used in this session have been drawn from the clinical domain, there is a real need for the bench based researcher to think in similar terms. In the development of ELISAs or PCR assays for the investigation of disease in research, there is a tendency to go with the best gels, be satisfied with the most convincing blots. But the simple fact that remains is that if results are not reliable, if associations are not demonstrably repeatable, then it is science that is the loser.

Alice had it easy.

REFERENCES

Abramson J.H., (1994) In: *Making sense of data*. Oxford University Press, Oxford.

Cohen, N.D. and Mackay, R.B., (1997) Interpreting immunoblot testing for cerebrospinal fluid for Equine Protozoal Myeloencephalitis. *Compendium on Continuing Education for the Practicing Veterinarian*, **19**, 1176-1181.

Good, (1995) When batterer turns murderer. *Nature*, **375** 541.

Kerr, (1989) In: *Veterinary Laboratory Medicine. Clinical Biochemistry and Haematology*. Blackwell Scientific Publications, Oxford.

Murphy, D., Reid, S.W.J. and Love, S. (1997) Modified oral glucose tolerance test as an indicator of small intestinal pathology of horses. *The Veterinary Record*, **140**, 342-343.

CLINICAL TRIALS: DESIGN AND ASSESSMENT

Stuart W.J. Reid

INTRODUCTION

This session is designed to provide an overview of clinical trials involving animals. The issues addressed will provide a grounding in the approach to assessment of a proposal or a report. This will allow a rational criticism of the clinical trial to be formulated in general terms. The module is not designed to provide an exhaustive list of experimental designs nor detailed instruction in the analysis of such trials. These constitute a science in their own right. The topics covered in these notes are based on "Clinical Trials" as set out in (Thrusfield 1995).

A clinical trial is a systematic study in a target species which is used to establish the efficacy of an intervention. The intervention may be therapeutic or prophylactic and may be a surgical, dietary, pharmacological or biological. The trial is designed to provide quantitative evidence in contrast to qualitative, subjective anecdotal assessment which is useless if the effect of the intervention is small.

In order to provide effective and safe products, regulatory bodies have required that product claims be substantiated by evidence based on competent analysis. Whilst these regulations and guidelines have been instituted for industry, they are essentially borne out of good scientific practice, that is,

- 1) Hypothesis formulation
 - based upon a sound knowledge of the literature
- 2) Hypothesis testing
 - by appropriate experimental design
 - and data analysis

Clinical trials take the form of controlled experiments. Comparing a group of animals before and after treatment provides some evidence but in an uncontrolled fashion: it is not possible to say that any perceived improvement would not have occurred anyway. It is best if a controlled clinical trial compares the treatment group with a concurrent group, not a historical one. This comparison or control group may receive another treatment, no treatment or a placebo.

In order to control bias in a study (differential assignment or management of the groups), randomisation is used (see below). The randomised controlled clinical trial is the gold standard. Blinding is another means of reducing bias. This may be single or double, but may not always be possible.

In other circumstances, trials may be conducted "in the field"

The essential ingredients of a clinical trial are:

- protocol
- primary hypothesis
- primary end point
- (secondary hypotheses and end points if appropriate)
- a decision on measure of efficacy
- definition of experimental units
- admission criteria
 - definition of condition
 - criteria for diagnosis
- exclusion criteria
- informed consent (if appropriate)

PROTOCOLS

As with any scientific experiment the requirement for clarity and structure is essential. The following is a list of key points that should be easily identifiable in any proposal or study report.

1. Studies must have a clear objective
2. Existing data on any product used must be provided. These include identity, formulation, dosage and toxicity.
3. A protocol must be approved before commencement
4. Studies must adhere to the protocol
5. Selection criteria of subjects must be supplied
6. Animal welfare and statistical considerations must be satisfied by the study design including numbers of animals used and randomisation procedures
7. Objective measures must be used to achieve study aims. Ideally and frequently these should be quantifiable parameters
8. Evidence that the study has been monitored and that complete and accurate original records are maintained
9. There must be a provision for the immediate reporting of adverse reaction
10. All reports must be factual, accurate and complete
11. receipt and storage of test substances must be documented

Once the scrutineer is satisfied that the points above have been adequately addressed, A more detailed appraisal of the protocol should be undertaken. It is helpful if all submissions adhere to a standardised format.

Objective

Introduction/background rationale

Animals

- species
- breed
- source
- sex
- age
- weight
- number
- identification

Study schedule

- start
- finish
- duration
- date of report

Facilities and personnel

Materials and Methods

- Inclusion/exclusion criteria
- Housing
- Management
- Feed and water
- Formulation
- Treatment groups / administration
- Allocation

Records/Observations

- Clinical measurements
- Sample collection and processing
- Health and adverse reaction monitoring procedures
- Other medication
- Deaths

Animal disposition

Drug control/accountability

Records (pro forma wherever possible)

Data handling and analysis

Final Report and archiving

EXPERIMENTAL UNIT

The experimental unit is the smallest independent unit to which the treatment is randomly allocated. It may be the individual animal or it may be a batch, pen or house. The experimental population is the population in which the trial is conducted. This population should be the target population. This will be discussed in more detail under validity.

RANDOMISATION**Simple randomisation**

This may be achieved by the tossing of a coin when there are two experimental groups but it is more common to use a source of random numbers. Note that randomisation should be performed after eligible subjects have been identified.

Block randomisation

When conducting a small trial, simple randomisation can lead to uneven distribution of animals among groups. This can be overcome by the use of block or restricted randomisation. Randomisation is limited to blocks of units ensuring equal numbers of animals in each treatment group.

Stratification or matching

During the randomisation process animals may be matched on one or more characteristics e.g. sex, age, weight. By doing this matching and then allocating the pairs randomly to treatment or control groups, bias due to the matched factors is avoided. Note that it is not possible to investigate the effect of factor by which matching or stratification has been performed.

TRIAL DESIGN

There are essentially three main designs

1. Standard
2. Cross over
3. Sequential

What follows are descriptions of these followed later in the text by schematic/tabulated examples of variations on the theme.

Standard trials

These involve random allocation of animals to treatment groups, the treatment applied and the groups followed for a predetermined time. The analytical methods used will depend on the response variable

Cross over trials

Animals receive more than one treatment, one after another. In this case the responses are paired. Care must be taken if it is suspected or known that there is a carry-over effect from the treatment. These designs are useful where the number of animals is limited.

Sequential trials

These are complicated designs and require careful attention to be paid to sample size. In sequential trials, analyses are repeated as data accumulates. This leads to an overall increase in the Type I error and necessitates the reassessment of what we consider as our level of significance. These designs are best avoided and will not be considered further.

DESIGN SENSITIVITY AND VALIDITY

Sensitivity is the likelihood that an effect, if present, will be detected.

Validity is the likelihood that the effect detected is the effect of interest. There are several types of validity that we must consider.

Internal validity

This relates to the legitimacy of attribution of effect differences between the treatment and control groups of the experimental population. Randomisation usually ensures internal validity.

Were the animals randomly assigned to treated and control groups?

Were the management conditions similar?

Construct validity

Does a reduction in CPK mean the animal is fitter?

Does an improvement in general demeanour mean the animal feels better?

External validity

This relates to the generalizability of results. If the experimental population is representative of the target population, unbiased inferences may be made regarding the performance of the test product in the general population. external validity is most likely to be achieved in field trials.

Are the study animals representative of the target population?

Will the drug be effective for animals on a different nutritional plane?

Statistical conclusion validity

Has the experimental design and choice of statistical test got sufficient power so that if the treatment did have an effect, the treatment versus control difference will be statistically significant?

PROBLEMS WITH CLINICAL TRIALS

Veterinary trials and experiments are often plagued by undue attention to design, which makes the cost of product development unnecessarily expensive and the trial ineffective. Some commonly recurring design faults in veterinary trials are:

1. **The sample size is too small** (or what to do if you don't want significance!)

This occurs frequently. The temptation to save on costs of experimental material and the effort involved in data collection often leads to experimental designs which have insufficient power. It is a false economy as in many cases it may be better not to undertake the trial at all.

2. **The sample size is too large** (or what to do to get statistical significance and not biological significance!)

This is not uncommon. In general the more data collected the more powerful the method of analysis becomes with the result that small orders of difference will be detected by a statistical test. Such differences may not be biologically important and so they are misleading. Fortunately, in recent years, most product development has been market-driven. Consequently, market forces often determine what order of difference must be achieved by a new product if it is to be profitable.

3. **There is a placebo effect** (or what to do if you want to waste time and money!)

This is a common phenomenon in trials where there is a subjective assessment of the effects of a compound. If the placebo is effective in improving the condition of 50% of subjects (not unusual) and the treatment has a 60% improvement rate, then a very large number of data from each group (at least 200 animals in each group) will be required to have a reasonable chance of establishing significant differences between groups.

Design faults 1, 2 and 3 can easily be avoided by using information from past studies to calculate the appropriate number of data to collect in a prospective trial.

4. **There are too many statistical tests** (or how to get at least one significant result and prove you haven't been wasting your time!)

In studies where there are more than two treatment groups, testing each group against each other group can lead to several statistical tests. Alternatively, when data are collected from several groups at different points in time there is a temptation to carry out a statistical test using the data from each time point. These actions increase the number of statistical tests used and increase the chances of a test falsely indicating a significant difference. *A basic principle of data analysis is to carry out as few statistical tests as possible.* By using an appropriate single test, such as a multiple range test or analysis of variance for repeated measures, it is often possible to avoid unnecessary statistical testing.

5. **Animals are not randomly assigned to treatments** (Or how to prove what you want, before you collect your data!)

Experimental designs require the randomisation of subjects to treatments. In the case of weight studies, the randomisation would normally be carried out using the initial weights of the animals. This ensures that treatment groups have animals of comparable weight at the start of the study. Failure to randomise can invalidate analysis of data collected in a trial.

6. **Introduction of bias and confounding** (or how to design your trial to prove what you want!)

Bias and confounding often arise unwittingly. Bias can arise as a result of data being omitted after the trial has been completed whereas confounding may arise from managing animals in treatment groups under different conditions. Proper attention to how the trial is to be conducted and the choice of experimental design can avoid bias and confounding.

7. **The investigator insists on using a multivariate analysis** (or how to make a simple problem difficult and impress your colleagues!)

The analysis of trials should be kept simple and the use of complicated statistical techniques avoided.

SOME COMMON EXPERIMENTAL DESIGNS

Two sample t test (& Mann Whitney)

Animal No.	Control	Animal No.	Treatment
1	x	6	x
2	x	7	x
3	x	8	x
4	x	9	x
5	x	10	x

treatment

Paired t test (& Wilcoxon)

Animal No.	Control	Animal No.	Treatment
1	x	1	x
2	x	2	x
3	x	3	x
4	x	4	x
5	x	5	x

treatment taking account of animals

Completely randomised design

Animal No.	Control	Animal No.	Treat A	Animal No.	Treat B
1	x	6	x	11	x
2	x	7	x	12	x
3	x	8	x	13	x
4	x	9	x	14	x
5	x	10	x	15	x

treatment

Randomised block design

Animal No.	Control	Animal No.	Treat A	Animal No.	Treat B
1	x	1	x	1	x
2	x	2	x	2	x
3	x	3	x	3	x
4	x	4	x	4	x
5	x	5	x	5	x

treatment taking account of animals

often crossed over with washout

Nested Design

Control			Treat A			Treat B		
Litter 1	Litter 2	Litter 3	Litter 4	Litter 5	Litter 6	Litter 7	Litter 8	Litter 9
x	x	x	x	x	x	x	x	x
x	x	x	x	x	x	x	x	x
x	x	x	x		x	x	x	
x	x	x	x		x		x	
x		x			x		x	
x					x			

treatment taking account of litter

useful for house studies

Two Factor Design

	Control			Treat A			Treat B		
Diet 1	x	x	x	x	x	x	x	x	x
Diet 2	x	x	x	x	x	x	x	x	x
Diet 3	x	x	x	x	x	x	x	x	x
Diet 4	x	x	x	x	x	x	x	x	x

treatment, diet, treatment*diet
useful for house studies

Two factor with repeated measures

	Control			Treat A			Treat B		
Animal	1	2	3	4	5	6	7	8	9
Time 1	x	x	x	x	x	x	x	x	x
Time 2	x	x	x	x	x	x	x	x	x
Time 3	x	x	x	x	x	x	x	x	x
Time 4	x	x	x	x	x	x	x	x	x

Treatment, time, animal, treatment*time

Two factor cross over with repeated measures

	Control			Treat A			Treat B		
Animal	1	2	3	1	2	3	1	2	3
Time 1	x	x	x	x	x	x	x	x	x
Time 2	x	x	x	x	x	x	x	x	x
Time 3	x	x	x	x	x	x	x	x	x
Time 4	x	x	x	x	x	x	x	x	x

Treatment, time, animal, treatment*time

Note that these designs are difficult enough to perform meaningful analysis (assuming adequate numbers, randomisation etc.) If a different design is proposed on presented, unless there are grounds for assuming that the design was carefully planned and executed, the chances are that the analysis will be difficult, impossible or a salvage job.

**THE RIRDC EQUINE INDUSTRY PROGRAMME
THE ROLE FOR EPIDEMIOLOGICAL RESEARCH
Stuart W.J. Reid**

1. BACKGROUND

Based on the vision and mission statement of the R&D plan for 1996-2001, the Australian equine industry, in concert with RIRDC, has a clear commitment to the health and welfare of Australian horses and the profitability of the industry in a global market place. As part of the mission, the eight specific research programs have played a pivotal role in highlighting areas of need, identified through extensive consultation with the industry. From the perspective of disease control, the implication is that a population-based approach is necessary as it is obvious that widespread or potentially widespread disease will have the greatest overall impact on economic and welfare standards. Epidemiology, which is the study of disease in populations and the quantitative assessment of the factors that determine its occurrence, distribution and severity, therefore has a central role to play in the R&D program. This report summarises the potential for epidemiological research in the existing Five Year R&D Plan, comments on the outcome of the RIRDC Epidemiology Workshop, and addresses some future issues not specifically mentioned in the Five Year Plan.

2. EPIDEMIOLOGY

2.1. What is on offer?

Advances in the molecular sciences have led to improved methods for detecting, treating and preventing disease but only relatively recently has the need to evaluate the efficacy of diagnostic, therapeutic and prophylactic intervention become apparent. Medical research now acknowledges that risk assessment and quantitative appraisal are pivotal in the implementation of advanced health care systems. Epidemiology, biostatistics, decision analysis and health economics are the “evaluative biological sciences” necessary for such investigations. With particular reference to the equine industry, in order to maximise the benefits to health and welfare of the thoroughbred by the most efficient means, quantitative assessment of available data from currently employed strategies must be performed. Whilst animal welfare is paramount, intervention at all levels must be carried out in the most cost-effective manner, if the industry is to prosper. Collaborative links with practitioners and trainers, in projects initiated by the industry, can ensure significance of findings to the field situation.

2.2. Types of Study

The application of epidemiological methods to equine disease problems can take several forms. These range from basic survey methods, through longitudinal studies to the large-scale intervention studies. Each of these methods has advantages and disadvantages and these in general relate to the economic cost of the study and the utility of the information that becomes available.

2.2.1. *Observational epidemiology*

- Cross-sectional studies; basic surveys. Cheap and of limited use; a good starting point.
- Case control studies; comparison of diseased and healthy animals and the factors to which they have been exposed. Cheap and useful for identifying

potential causes of disease occurrence. Unable to prove causal relationships. Can be performed quite quickly if existing data available.

- Follow up studies; comparison of two groups of animals exposed to different conditions or agents. Wait to see which ones become diseased. Expensive and time consuming. Causal relationships may be identified.

2.2.2 *Experimental epidemiology*

- Clinical trials; deliberate intervention with new drugs vaccines etc. under laboratory-type conditions.
- Field trials; a term applied to later stage clinical trials conducted fully in the field involving field cases.

2.2.3. *Other activities*

These include activities that do not fit into any one of the categories above;

- Modelling; simulating disease occurrence with mathematical, statistical and computing techniques.
- Outbreak investigation; Emergency epidemic investigation and intervention
- Surveillance; On going reporting of disease occurrence

2.3. **Advantages**

From the initial identification of the diseased animal through diagnostic work up, clinical and pathological examination and testing, tracing of contacts, identification of source, assessing economic impact and predicting future outbreaks it is unlikely that one individual or even one discipline will be exclusively involved. Epidemiology perhaps is the underlying or unifying science that brings together the experts from all of the appropriate fields.

There are some obvious advantages to observational epidemiology;

- ***Non-invasive***: Projects can make use of existing information or study the incidence and distribution of naturally occurring cases of disease. No animal experimentation need be involved.
- ***Evaluative***: Assessment of efficacy, efficiency and cost benefits of management techniques, therapy and current knowledge.
- ***Multidisciplinary***: The approach builds on strengths and brings together many interested parties. Projects can often be initiated by the equine industry or be conducted in collaboration with equine practitioners.
- ***Relevant***: Studies conceived through industry consultation addressing industry needs and conducted in the field are likely to be relevant to the industry and have outcomes applicable to field situations.

3. **THE FIVE YEAR PROGRAMME**

It is clear that for each of the eight programs major advances have been made.

3.1. **Program 1. Information and the Horse Industry**

Key to this program has been the development of Web based technologies and the impact of this communication system is going to continue to grow in importance. This program complements the application of epidemiology both in terms of flow of data and information to the researcher but also and arguably more importantly, the flow of

information and knowledge back to the industry. Program 1 therefore must be regarded as being of major significance in terms of **Technology Transfer**.

3.2. Program 2. Industry Development and Development of Research Infrastructure in Australia.

This program will continue to be of importance given the limited size of the industry and particularly the size of the equine research community worldwide. Attention will increasingly be paid to collaborative projects and these will only be successful if there is a well-developed infrastructure both within the industry and between the industry and research communities, nationally and internationally. This awareness and commitment must focus on international disease standards.

3.3. Program 3. Respiratory diseases

These diseases will continue to be a problem on a global basis and with the international movement of horses and the potential economic impact on equine health, the industry in Australia will continue its active support of this research. Key questions that remain relate to the prevalence and cause of endemic respiratory disease and the risks of importing new diseases must be evaluated on a continual basis. Opportunities exist to evaluate respiratory disease on an international basis and it is likely that a concerted action will lead to more rapid advances. In addition, knowledge and attitude to “bleeders” may provide a useful focus for collaboration.

3.4. Program 4. Nutrition research

In many respects this is a complex area and the interaction of the genetic, exercise and feeding factors make investigation appropriate using epidemiological techniques. However, optimal feeding of production animals has benefited from large studies and a restricted number of outcome requirements. It is likely that similar large-scale studies will also be required for equine nutrition.

3.5 Program 5. Lameness and Limb Injury

Recent work has provided enormous advances in this area but it is less clear on a population wide basis how training methods affect disease incidence and prevalence. From “accepted” conditions such as shin soreness through more recently acknowledged problems like DOD, to the fatal injuries, a need remains for baseline rates of occurrence in different horse populations. If these are not known, how then can interventions be assessed other than against subjective opinion? New approaches are required for areas where confidentiality and economic impact are a consideration.

3.6 Program 6. Exercise and Metabolic disease

Experimental studies have been widely applied to many of these diseases to good effect but the application of population studies is required for conditions such as tying up where existing *tenets* are now being called into question. Once again base-line incidence and prevalence data will be a critical first step.

3.7 Program 7. Reproductive disease

Manipulation of the reproductive cycle underpins the supply of animals to the racing industry and yet the levels of wastage and failure remain high in some areas. The impact of conditions in early pregnancy on subsequent performance remains another area for consideration. From an epidemiological point of view the use of artificial breeding technologies provides a useful control method for certain disease conditions.

3.8 Program 8. Improving racetracks and working surfaces

The long-term strategy of improved surfaces for the working animal is an important objective. However, the necessity to evaluate injury rates on these different surfaces controlling for type of work or training method should be important primary objectives. In this respect there will be substantial interaction with Program 5.

4. RIRDC EPIDEMIOLOGY WORKSHOP

4.1. Lecture Sessions

The presentation sessions during the workshop covered the most significant techniques currently being used to address animal health at the population level. What was apparent was the common interest among all the participants and that there was a need to disseminate the fact that epidemiology constitutes more than biostatistics, a frequently encountered misconception worldwide. The relevance of the methodological sessions was reinforced by selected case studies currently underway in Australia, some funded by the RIRDC. The interaction during the sessions raised several key themes including

- the validity of much previous research
- the necessity for appropriate experimental design and data analysis
- the requirement for adequate quantitative training for researchers
- the need to address disease problems with a multidisciplinary approach
- the necessity to focus on industry problems
- the need to focus on modifiable risks for disease
- the requirement to ensure adequate and appropriate technology transfer

4.2. Brainstorming session

Led by Professor Rose and Ms Lublin, the brainstorming session generated over 50 issues the participants at the workshop viewed as being of importance to the equine industry in Australia. This list was then condensed through an interactive clustering process that resulted in eight main themes viz.

1. Respiratory disease
2. Communication, training and education
3. Population and disease surveillance
4. Musculoskeletal problems of adults and foals
5. Research funding and collaboration
6. Reproductive rates, disease and foal health
7. Feeding
8. Therapeutics

It is indeed heartening that these major themes overlap substantially with those outlined in the individual program topics listed above. However, it was also notable that the participants, when addressing the issues of

- Funding sources
- Expertise
- Infrastructure

were unanimous about the requirement for industry (pharmaceutical, retail and racing) commitment to funding, for multidisciplinary collaboration and the need for regional co-ordination.

The most significant additions to the program topics were

1. The desire to implement national disease surveillance, quarantine control and establish population demographic characteristics

2. Education and training at all levels of equine owner

In both of these topics it is likely that information technology will play an important part.

5. The Future

The program initiatives contained in the R&D plan are evidently proceeding well. Epidemiology, as one of the underlying sciences, has much to offer in that the majority of the diseases that compromise health, welfare or economic performance are industry wide, yet their effects are still relatively poorly quantified. If Australia is to maintain its equine industry at the forefront of innovation, a clear understanding of the measurable constraints on the industry is required. Measurement is a first step towards understanding the actions that can lead to beneficial change on a cost-effective basis. RIRDC is in the position to support these activities but cannot adequately fund all necessary projects. The support of “problem-solving research for industry” that will see the industry move forward is of utmost importance and one must assume that, in the current climate, central sources will continue to decrease in real, if not in absolute, terms. Australia is unique in its commitment to the equine industry. The task must now be to convince those that stand to benefit most from advances in disease control, that cost-effective research is a staple commodity and not a luxury.

In Europe, there are moves to identify horses by microchip and to record certain therapeutic drug administration on an individual animal basis. The benefits of such a program go far beyond control of substances in the human food chain as, with individual identifiers, the potential for data gathering and disease control at the population level will be second to none. Australia has an enviable quarantine status and it is my belief that through commitment to continuing innovation Australia can continue to lead the way. Epidemiology will be a useful tool in achieving this goal.

**WORKSHOP ON MAJOR PROBLEMS FACING THE EQUINE
POPULATION OF AUSTRALIA**
Prioritising Equine R&D Issues

During the Epidemiology Workshop, a planning exercise was conducted to address the key issues of problems facing the equine population of Australia. Professor Reuben Rose introduced Dr. Jackie Lublin, who facilitated the afternoon practical workshop session.

The session commenced with each participant asked to list what they perceived to be the major issues/diseases facing the equine population of Australia. Each participant then read an issue from their list until all choices had been documented on one list (this process lent itself to each participant listing approximately two choices each).

The initial list produced 53 issues and each participant then was asked to list the eight most important of these issues which allowed determination of the priority of each of the 53 issues. Further refinement was possible by clustering 'like' issues together, resulting in eight key issues.

The participants then broke into small groups, each responsible for one particular issue. Their task was to list what would be needed to effectively tackle this issue/disease in terms of:

Money
Expertise
Other

The 53 identified issues are listed with the number of votes from participants reflecting the perceived importance of each of these:

1. Respiratory disease management and prevention 18
2. Developmental orthopedic disease 10
3. Low reproductive rates in the thoroughbred industry 6
4. Exercise induced pulmonary hemorrhage 3
5. Musculo-skeletal disorders 10
6. Communication between industry and researchers 10
7. Client education 7
8. Welfare issues associated with eventing, steeplechasing hunting etc. 4
9. Overuse of antibiotics 6
10. Transport associated wastage 2
11. Role of environmental and management aspects with disease 3
12. Stress – role in modifying immunity 1
13. Wastage due to shin soreness in two year olds 4
14. Objective evaluation of racing and training surfaces 4
15. Medication and medication control and the given increasing sensitivity of measurements 1
16. Quarantine and quarantine control with increased horse movements 6
17. Lack of equine demographic data 8
18. The challenge of disease from the 'north' 2

APPENDIX II

19. Feeding for optimum growth and performance 3
20. Increase community concern about welfare issues 5
21. Cyathostome burdens in horses 1
22. Shoeing and foot care 1
23. Foal health issues such as rhodococcus equi 5
24. Control of performance modifying drugs that are difficult to detect 3
25. Better defining problems in each of the major breed/use groups 6
26. Training and skills for the people in the industry 5
27. Identification of optimum training programs for different horse groups and uses 2
28. Parasite control 3
29. Environmental/welfare issues for urban horses 2
30. Impact of offshore racing – eg. Asia 1
31. Standardizing recording of clinical information 1
32. National diagnostic standards 2
33. Adequate funding of research 10
34. Stimulating research collaboration 4
35. Feed practices and digestive physiology 4
36. Links between practitioners and researchers and institutions 4
37. Loss of genetic diversity through overuse of particular stallions 1
38. Benchmarking performance 4
39. Joint disease in two year olds 0
40. Artificial breeding disease 1
41. Neonatal and adult horse diarrhoea 0
42. Sensitivity and specificity of current tests 1
43. Solving dilemma of industry self funding for diseases 2
44. Development of export trade 1
45. Economic and social issues identification relating to the horse industry 2
46. Drench resistance 2
47. Behavioral problems 0
48. Improved procedures to characterize novel diseases 2
49. National disease (infectious) surveillance 13
50. Vaccine development 3
51. Improved understanding of bizarre diseases e.g. Ross river, stringhalt 2
52. Alternative therapies 1
53. Sexually transmitted?? reproductive diseases 8

Grouping the Major Issues

Grouping together of the key issues identified and their importance was undertaken by the group and resulted in the following priority list.

1. Respiratory disease. Total votes = 26
2. Demography/surveillance/quarantine – population and disease surveillance. Total votes = 27
3. Musculoskeletal problems of adults and foals. Total votes = 24
4. Communication, training and education. Total votes = 22
5. Research funding and coordination, collaboration. Total votes = 20
6. Reproductive rates and disease and foal health. Total votes = 19
7. Feeding. Total votes = 9
8. Therapeutics. Total Votes = 13

The workshop then divided into 4 groups to consider further some of the major issues relating to the top 4 priorities that had been identified. It was noted that these priorities reflected overall the major thrust of the 5 year plan but that the key issue of population and disease surveillance had been identified as an additional area that previously had not been considered.

RESPIRATORY DISEASE

Key Issues

- Management and prevention
- Exercise induced pulmonary haemorrhage
- Transport associated problems
- Environmental and management aspects

What will be needed to effectively tackle this issue/disease in terms of:

MoneySources

- ◆ RIRDC
- ◆ Racing Victoria Equine Research Fund
- ◆ ARC – Collaborative grants
- ◆ Commercial
 - ◆ CSL }
 - ◆ Websters } Vaccines
- ◆ Boehringer
- ◆ Hoffman – LaRoche?
- ◆ International funding

Expertise (in what areas? Is it available in Australia? If not, then what?)

Exercise Physiology

Camden – Rose, Hodgson, Evans
Queensland – Thornton

Bacteriology

Sydney – Daria Love

Virology

Melbourne – Studdert
Sydney – Hodgson

Rhodococcus

Angela Begg
Glen Browning
Mary Barton

Other (infrastructure, political will, corporate vision, willingness to change, etc)

- ◆ Coordination of Funding/Projects – RIRDC
- ◆ Communication Between Research Groups to:
 - ◆ Decrease overlap of research
 - ◆ Increase efficiency of use of infrastructure, personnel, and animals
- ◆ Collaboration Between groups
- ◆ Look to other sources of infrastructure e.g.: human labs, medical researchers

COMMUNICATION TRAINING & EDUCATION

Key Issues

- Communication between industry and researchers
- Client Education
- Training & skills enhancement for people

What will be needed to effectively tackle this issue/disease in terms of:

Money (from where)

- ◆ Industry
- ◆ RIRDC
- ◆ TAB

Expertise (in what areas? Is it available in Australia? If not, then what?)

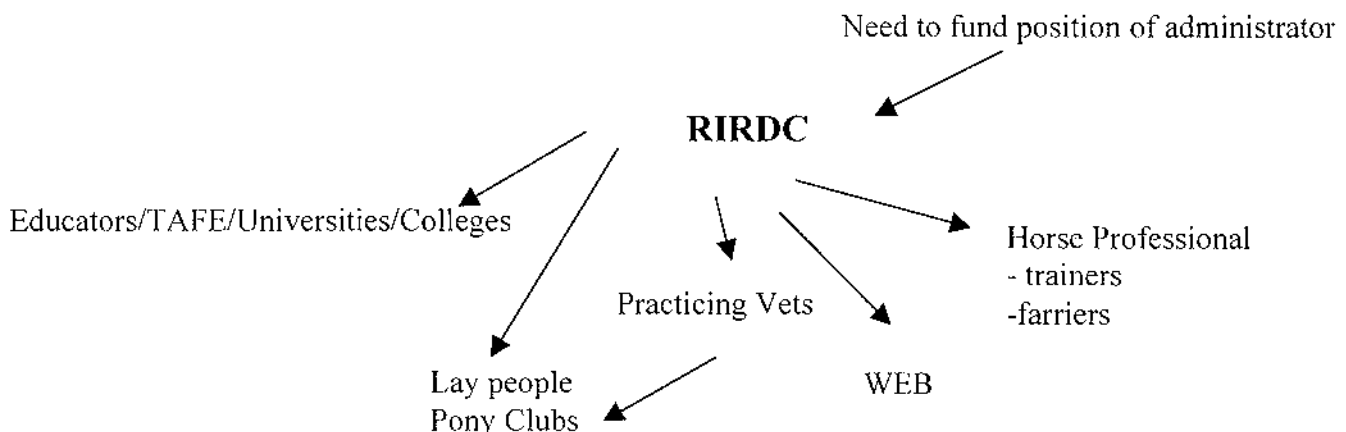
- ◆ Researchers
- ◆ Journalists/Magazines/Websites
- ◆ Building on existing RIRDC programs

Other (infrastructure, political will, corporate vision, willingness to change, etc)

Existing Infrastructures:

- ◆ Pony Clubs
- ◆ Racing Authorities
- ◆ Special Interest Groups – TBA

There is a need for a central compositor/distributor of information



POPULATION & DISEASE SURVEILLANCE

Key Issues

- National disease surveillance
- Quarantine control
- Lack of equine demographic data

What will be needed to effectively tackle this issue/disease in terms of:

Money (from where)

- ◆ Sell concept/justification
- ◆ Industry & government matching dollar for dollar
- ◆ Identify further key points for levies

Expertise (in what areas? Is it available in Australia? If not, then what?)

- ◆ Lacking state based coordinators
- ◆ Epi-skills are available here
- ◆ Learn from other production industry groups e.g. Pigs
- ◆ Establish contact with other similar programs overseas e.g. MEMS, Equinella, 'Equiness'

Other (infrastructure, political will, corporate vision, willingness to change, etc)

- ◆ State based data collection
 - ◆ National body e.g. NAHIS
- ◆ Continued membership of AAHC
- ◆ Influential government and industry advocates needed but from where?
- ◆ Business Plan / ID surveillance needs
- ◆ In terms of willing to change – part of external environment – trade as one degree driving force.

MUSCULOSKELETAL PROBLEMS OF ADULTS AND FOALS

Key Issues

- DOD in foals
- MS disease
- Wastage due to shin soreness
- Joint disease in two year olds
- Shoeing and foot care

What will be needed to effectively tackle this issue/disease in terms of:

Money (from where)

- ◆ RIRDC
- ◆ State Governments

- ◆ Racing Authorities
- ◆ Breeders

Expertise (in what areas? Is it available in Australia? If not, then what?)

Developmental Orthopaedic Disease

Much preliminary work exists and ongoing cost to industry estimated – interest of breeders established

Proposal

Large study of cohort of foals

Severity, incidence plus other problems of foals

Nested case control studies (simultaneous collection of data about risk factors)

Control of DOD could require modification in presentation of yearlings. Is there political will to change?

- ◆ Epidemiologist
- ◆ Clinicians/Local Vets
- ◆ Cooperative/Interested Stud
- ◆ Staff

Other (infrastructure, political will, corporate vision, willingness to change, etc)

- ◆ Knowledge of existing records
- ◆ Equipment? – Scales – willingness to use them
- ◆ Similar study with older or same cohort study re: shin soreness
- ◆ Follow original DOD cohort to start of racing
- ◆ Willingness to change?

RESEARCH FUNDING AND COLLABORATION

Key Issues

Increased research funding

Stimulating research collaboration

Links between practitioners, researchers and institutions

Industry self funding for disease

Reproductive Rates, Disease and Foal Health

Low reproductive rates

Sexually transmitted diseases

Foal health issues

Feeding

Feeding for optimum performance

Feeding practices and digestive physiology

Therapeutics

Overuse of antibiotics

Medication control given increased sensitivity of testing

Parasite control

Drench resistance

Alternative therapies

Vaccine development