Statistical Methods in
Veterinary Epidemiology

Introduction to the Course

Welcome to the course

Welcome to *Statistical Methods in Veterinary Epidemiology*. This is a core course in the Veterinary Epidemiology and Public Health programme of the Royal Veterinary College.

The course is designed to provide you with the key statistical knowledge, understanding and skills you will need to analyse and interpret data from common forms of epidemiological studies that are conducted in veterinary science. The approach aims to enable you to gain an understanding of the concepts used in statistical analyses and modelling and will help you to choose and apply techniques appropriate for parameter estimation and hypothesis testing in selected situations.

You should be familiar with the basic concepts of statistics and epidemiology before undertaking this course.

What will you learn from this course?

By the end of this course you should be able to:

- describe the basic statistical measures and concepts underlying the analysis of epidemiological data
- use a comprehensive set of statistical methods suitable for a wide range of epidemiological situations
- select appropriate statistical techniques for the analysis of data from epidemiological studies
- identify specific issues relevant to case–control and cohort studies
- perform basic statistical modelling techniques
- investigate confounding and interaction in epidemiological data using both stratified analyses and statistical modelling methods
- interpret the results of statistical procedures and draw appropriate conclusions.

Although this course has been designed to avoid unnecessary mathematical detail, a proper understanding of the methodology does require the use of some mathematical formulae; however, you will not need to learn all of the formulae by heart. No knowledge of calculus will be required.

You may find some of the material in the course quite challenging, but when you have completed it you should possess a comprehensive knowledge of statistical analyses and modelling techniques at a level desired by all epidemiologists.

Course structure

The course consists of three modules comprising 12 computer-assisted learning (CAL) sessions and a Workbook containing seven practical computer sessions using the statistical package R. You are advised to work through the CAL sessions and their associated Workbook sessions in the sequence in which they appear in the three modules.

The CAL sessions

The CAL sessions are all on a CD-ROM supplied in your study pack. These files have been designed to run on the web browser of your computer.

The study sessions are made up of easily navigable sections and subsections, providing the information you need for SMVE in a straightforward and digestible format. There are interactive sections where you will be asked to answer various types of questions; some involve calculation and others require you to think. You will also be directed to read sections of the literature provided with this course.

The CAL sessions are grouped into three modules:

Module 1 *Introduction to Multivariable Analysis of Epidemiological Data*

In the first module of the course the focus is on multivariable epidemiological analysis. In Sessions 1–6 you will go through a series of epidemiological concepts that are essential for the later application of specific multivariable techniques, the focus of Modules 2 and 3. This module looks at the use of measures of effect and the analysis of cohort and case–control studies, and introduces the concepts of likelihood and the analysis of time to event data.

Module 2 *Logistic Regression*

The second module deals with logistic regression, which is used to model binary outcome data. The three sessions explain how to deal with interaction in a logistic regression model and how to use and interpret logistic regression for quantitative exposures.

Module 3 *Regression Models for Time to Event Data*

Following the introduction to multivariable epidemiological analysis and statistical modelling in Module 1, and the study of logistic regression to model binary outcome data in Module 2, this final module allows you to study regression models for time to event data: Poisson regression and Cox regression.

The Workbook sessions

Seven of the CAL sessions are accompanied by practical sessions which are set out in the printed Workbook in your study pack. These practicals use the software package R, which you will download from the internet.

The Workbook sessions give you the opportunity to gain experience in performing your own analyses and are designed to reinforce your understanding of the techniques you have learned. You should work through each of these practicals directly after completing the corresponding CAL session.

Each practical session consists of a tutorial-type exercise that guides you through the analyses. In these tutorials you will be instructed as to which commands to use and provided with interpretation of the output – being able to use R to produce output is important, but the interpretation of the output is just as important, if not more so. At the end of each tutorial there is a review exercise where you will practise on your own and apply the commands used in the tutorial. The solutions to these review exercises are also given in the Workbook.

Most of the example datasets used in the practical sessions come from studies in which RVC staff have been closely involved. Although these datasets may not be appropriate for your particular field of work, you should find that they provide meaningful examples of specific study designs to which the methods are applied. A description of each dataset is given in the Workbook.

Once you have completed the CAL and Workbook sessions you should have the necessary knowledge and understanding to meet the learning objectives listed at the beginning of this Introduction. In order to achieve them, you should work through all of the activities, carefully checking your results against those given and investigating the reasons for any differences. Unless otherwise instructed in the activity, don't be tempted to move on to the feedback until you have completed the calculations or answered the question.

Computer and resources

To complete the work in this course you will require a computer with access to the internet.

In your study pack is a CD-ROM for the course containing two folders, the contents of which you will load onto your computer:

- 'SMVE CAL sessions'
- 'Tutor-marked assignments: Data files and data documents'.

You will need a basic scientific calculator with a natural log function (ln).

You have been provided with two textbooks for this course:

- *Multivariable Analysis: A practical guide for clinicians* by M.H. Katz (2006)
- *Veterinary Epidemiologic Research* by I. Dohoo, W. Martin and H. Stryhn (2009).

Katz (2006) focuses on quantitative epidemiological methods and shows you the proper application of these methods and the sound interpretation of analysis results. Dohoo *et al.* (2009) provides a good introduction to multivariable analysis and statistical modelling.

You might find it useful to have access to a basic statistical textbook in case you need to improve your understanding of some concepts referred to in the course.

All the other reading material you require is reproduced in the Reader, which forms part of the study pack.

Study time

The notional study time for this course is 240 hours, although it may take more or less time depending on your familiarity with the material and your computer competence. As a general guide it should take you about 120 hours to complete the study sessions (CAL sessions and practical sessions in R); you should spend about 40 hours on background and essential reading, about 20 hours on the tutor-marked assignments, and the remainder of the time on revising for the examination.

Assessment

Your work for this course will be assessed by means of a three-hour unseen written examination paper. In addition, you are required to complete and submit at least one tutor-marked assignment (TMA) for assessment. If you submit more than one – and you may submit up to three – your best TMA will be used in the calculation of your final mark. Full information on how to approach and submit TMAs is provided in your *Student Handbook* and in the assignments themselves. You should bear in mind that your TMA will count for 20% of your final mark for the course.

Preparing for the computer sessions

The electronic parts of the course require some straightforward preliminaries.

Loading the CAL sessions

The CAL sessions run on the web browser of your computer. The format is quite simple; you should become familiar with the layout quickly and find it very easy to use.

To open the course documents, insert the CD-ROM supplied for SMVE into your disk drive, picture side up.

If the contents of the disk are not automatically displayed, go to the 'Start' menu, click on 'My Computer' and then select the appropriate drive, such as DVD/CD-RW Drive (D:).

Once you have opened the CD, double-click on the folder 'SMVE CAL sessions' and then double-click on the appropriate session folder.

In order to open a session, double-click on 'index.htm'. The session will then appear in your web browser. You may be given the option of allowing blocked content (depending on your security settings); you should do this to enable the interactive parts of the course.

Navigating

Each session is divided into sections and subsections; the layout is very simple and user-friendly, and you won't get lost or skip a section unintentionally.

Each section of the session is labelled with a capital letter. You can access each section (or subsection) simply by clicking on its title. The content will be similar to the page shown in Figure 1.

Within a session you can navigate back and forth using the yellow arrows or click on the yellow circle to return to the contents page (alternatively, there is a functional contents list with hyperlinks on the left-hand side of the page).

Within the sessions you will find hyperlinks underlined in blue. These might be questions, in which case the solution will appear by clicking on the link, or they might provide additional information or diagrams. You will also find 'gap fill' and multiple-choice questions embedded in the sessions; these should be self-explanatory.

These instructions, and a preview of the content of each of the CAL sessions, can also be found in the file called 'SMVE synopsis' on the CD-ROM.

Downloading and installing the software

The Workbook sessions use the program R, a free, downloadable statistical software package.

In your browser, go to http://www.r-project.org and then click on the link **CRAN mirror**. This brings up a list of URLs for different countries. Click on the one nearest to you. Under the heading 'Download and install R', click on the appropriate operating system. If you select 'Windows', you should then click on **base**. From here you can download R 2.7.2 (click on it and then run).

Further instructions to get you started on using R are given in Section 1 of the Workbook.

Figure 1 Example of a CAL session