EPISCOPE: Computer programs in veterinary epidemiology

K. Frankena, J. P. Noordhuizen, P. Willeberg, P. F. van Voorthuysen, J. O. Goelema

The Veterinary Record (1990) 126, 573-576

Veterinary epidemiology is a rapidly developing science. However, many veterinarians are unfamiliar with the relevant techniques because veterinary schools have not introduced biostatistics as a core element of their courses or adopted epidemiology as a specific discipline. EPISCOPE, the computer software presented in this paper, covers many epidemiological principles and calculations. It can assist both the teaching of epidemiology and the analysis of field data.

The increasing importance of veterinary epidemiology is reflected in a series of developments. Societies for veterinary epidemiology have been established in the United Kingdom, Denmark and recently in the Netherlands, Scandinavia and France. New textbooks specifically relating to veterinary epidemiology are now available (Thrusfield 1986, Martin and others 1987). The need for systematic epidemiological and economic analysis of integrated veterinary, herd-management and economic records has increased (Dijkhuizen 1989) because advice to the farmer must be based on scientifically justified arguments for production policies, and the interpretation of herd data is unsound without the application of sound statistical techniques.

The application of quantitative epidemiology is germane to the collection and analysis of data (Rothman 1986). However, many veterinarians are unfamiliar with the relevant techniques because veterinary schools have not introduced biostatistics as a core element of their courses or adopted epidemiology as a specific discipline. As a result, computer programs which assist the teaching of epidemiology and the analysis of data have a valuable role in the teaching of epidemiology and the analysis of field data.

EPISCOPE consists of four modules, each module comprising a number of programs. The modules deal with the evaluation of diagnostic tests, sample size calculations, the analysis of cohort and case control studies, and models. Each program is 'menu-driven' and has a page-like structure (Fig 1). The pages are ordered vertically and access to different pages is obtained by selecting the page on the menu line. Below the menu line, extra information (menuhelp) is given about the option highlighted. The RESULTS, INTRODUCTION and HELP sections may consist of more than one page.

The WELCOME page contains information about: the name (function) of the program; the required input parameters; the output parameters; the menu on the status line which shows the next possible action, and the menuhelp on the bottom line.
which gives extra information about the next possible action.

Page 2 is the INPUT OF DATA page. Some text or lines, eg. a 2 x 2 contingency table, will be present on this page. Values for parameters can be entered by choosing the ‘Edit values’ option. The spreadsheet will be recalculated automatically. The results of this calculation can now be monitored by choosing the ‘Results’ option. In general, there is no input capacity on these pages, but in some programs a confidence level might be changed, eg. from 95 per cent to 90 per cent.

The INTRODUCTION and HELP pages are below the results. These give information about the subject with which the program is dealing. In some programs, these pages are followed by SCRATCH pages, for example for intermediate results, or graph titles, which are not informative to the user.

EPISCOPE is started via a batch file that loads SUPERCALC and a ‘macro’. A macro is an additional program that minimises the number of keystrokes and commands which need to be performed by the user. Within EPISCOPE, macros control the menu, making it very user-friendly. The first macro calls for a starting file that contains a list of the four modules within EPISCOPE. Next, the programs within the chosen module are displayed and the program of choice is loaded by typing its number. The program is entered via the WELCOME page.

Data

After loading the program initial data are available. These data are derived from examples in the accompanying workbook, which contains both fictitious data and data from published papers. The workbook has been developed for teaching purposes but is also useful for a first acquaintance with EPISCOPE. The initial data can be overwritten by other data.

In some EPISCOPE programs the input consists of frequencies, eg, the number of diseased and exposed animals in case-control studies. EPISCOPE is not meant for the calculation of these frequencies because SUPERCALC was not developed as a database-oriented program. It is often easier to use a real database program for that purpose.

Statistics

The statistics relating to case-control and cohort studies are based on the formulae of Rothman (1986).

Examples

Two examples of EPISCOPE programs are presented. Program 1 concerns the calculation of sample sizes for surveys. Program 2 determines the odds ratios and confidence intervals in case-control studies.

Results

Example 1: Sample sizes for surveys

When designing a survey for estimating the prevalence (P) of a disease, the sample size needs to be determined (number of herds, number of animals per herd). The formula to calculate the minimal sample size depends on the allowable error (L), the standard deviation (sd) of the expected prevalence and the desired level of confidence (Snedecor and Cochran 1980). The sd equals P(1-P), because the prevalence P is a binomial proportion. Thus, to assess the sd, an advance estimate of the prevalence, is needed. The following formula is used to determine the sample size, n:

\[ n = \frac{P \cdot (1-P)}{L^2} \]

where \( t \) is the value of Student’s \( t \) for normally distributed data at a specified confidence level

\[ P = \text{the estimated prevalence} \]

and \( L = \text{the accepted error} \)

This formula shows that the minimal sample size, \( n \), is larger when the desired confidence level is higher (\( t \) increases), when the allowable error is smaller, or when the expected prevalence is closer to 0-50 (or 50 per cent). However, the calculated value of \( n \) is only approximate, and needs to be adjusted when the sample size exceeds 5 per cent (rule of thumb) of the total population (N):

\[ n \text{ (adjusted)} = n/(1+f) \]

where \( f = \text{the sampling fraction} \) \( n/N \)

This adjustment is often necessary when small populations are sampled.

This calculation can be done by hand or with a pocket calculator but will be time-consuming especially when it is necessary to evaluate the effect of changes in \( L \), \( t \) or the expected prevalence. The EPISCOPE program SSPROPOR (sample size to estimate a proportion) has been developed to make quick calculations. After SSPROPOR is loaded the WELCOME page will appear (Fig 2).

The two INTRODUCTION pages of this program explain the formulae that are used. The page for the INPUT OF DATA and RESULTS (a combined page), including the fictitious data, is shown in Fig 3.
Example 2: Case-control studies

In case-control studies, the relationship between exposure to a factor and the occurrence of disease is analysed. Thus, there are four categories of animals (or herds) in a 2 x 2 contingency table (Table 1).

| TABLE 1: General structure of a 2 x 2 contingency table for a case-control study |
|---------------------------------|--------|--------|--------|
| Exposed                        | No     | Totals |
| Case                           | A      | B      | A+B    |
| Control                        | C      | D      | C+D    |
| Totals                         | A+C    | B+D    | A+B+C+D |

A, B, C and D represent the numbers of animals in each category. The strength of association between exposure and disease is expressed by the odds ratio (OR) which is calculated as:

\[
OR = \frac{(A/B)}{(D/C)} = \frac{AD}{BC}
\]

When the odds ratio is greater than 1.0 there is a positive relationship between exposure to the factor and the disease, and the factor is thus a risk factor. When the odds ratio is less than 1.0 the factor is associated with a reduced risk of disease. Once the frequencies in each cell are known the calculation of the point estimate of the odds ratio is simple. However, what really is of interest is whether the odds ratio is significantly different from 1.0. A confidence interval can be constructed for the odds ratio, and when this interval does not include 1.0 exposure to the factor is significantly associated with the disease (either positively or negatively, depending upon the values of the lower and upper confidence limits). The interval can be calculated in several ways. Two approximate methods (logit and test-based) are used in this program.

**INPUT OF DATA (A, B, C, D in upper part)**

<table>
<thead>
<tr>
<th>Observed frequencies:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
</tr>
<tr>
<td>Case</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

**Expected frequency of:**
- exposed cases (A): 39.58
- non-exposed cases (B): 11.42
- exposed controls (C): 57.42
- non-exposed controls (D): 16.58

**RESULTS**

| Odds ratio = | 4.15 |
| Attributable proportion = | -68 |
| Attributable proportion among exposed = | -76 |

<table>
<thead>
<tr>
<th>Desired level of confidence (90, 95, 97.5, 99, 99.5): 90.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence limits Odds ratio:</td>
</tr>
<tr>
<td>Logarithmic approx.</td>
</tr>
<tr>
<td>Chi-square approx.</td>
</tr>
<tr>
<td>Message: limits are</td>
</tr>
</tbody>
</table>

**FIG 6: RESULT page of program oacaseco, using data from Schukken (1988)**

On the lower part of the INPUT OF DATA page the expected frequencies are calculated. If any of these frequencies is less than five more complex methods need to be used to evaluate the relationship between disease and exposure (Rothman 1986). For the data in this example the odds ratio is 4.15 (Fig 6) and the lower limit of the 90 per cent confidence interval is greater than 1, indicating that there is strong evidence of a positive relationship between the use of teat dip and coliform mastitis. According to these data, teat dip cannot be recommended to prevent coliform mastitis on farms with low cell counts. The program also calculates the attributable proportion (AP) (Fig 6), that is the proportion of cases due to exposure. The
TABLE 2: Program available in EPISCOPE.

Module 1: Diagnostic tests
1 Test agreement
2 Test evaluation

Module 2: Sample size calculations
3 Sample size for detection of disease
4 Sample size to estimate a proportion
5 Sample size to estimate a difference between 2 proportions
6 Sample size to estimate a difference between 2 means

Module 3: Observational-analytical studies
7 Case-control studies
8 Stratified case-control studies
9 Matched case-control studies
10 Cohort studies (cumulative incidence data)
11 Cohort studies (incidence rate data)
12 Stratified cohort studies (cumulative incidence data)
13 Stratified cohort studies (incidence rate data)

Module 4: Models
14 Classical Reed-Frost model

AP indicates the relative importance of exposure to the occurrence of disease. It is the proportion of cases that would not have occurred if the exposure factor had been absent from the population. This parameter reflects causality. It is calculated as (OR-1)/OR. In the coliform mastitis example, the AP = 0.68 which means that 68% of all cases were related to the use of teat dip. Moreover, farms using teat dip have a higher prevalence of disease.

The last program, 14, deals with the classical Reed-Frost model, which facilitates epidemiological calculations. These calculations are often necessary for a correct and easy interpretation of data, but in practice are often not performed because veterinarians are unfamiliar with the underlying epidemiological principles and formulae. For example, data are often presented as percentages, regardless of absolute numbers, and the 'analysis' may merely involve comparing these percentages. As a result false interpretations may easily occur, with detrimental consequences. EPISCOPE has been developed because of this lack of knowledge. The program can be used by people who are unfamiliar with PCs. Furthermore, a knowledge of SUPERCALC is not needed because of the macrocontrolled menus. The EPISCOPE 'manual' consists of only one page.

EPISCOPE has been developed from the prototype EPIDEMO which was the result of a joint European Community project between the State University of Utrecht and the Royal Veterinary and Agricultural University, Copenhagen (Voorthuysen and others 1988). EPISCOPE is meant as a dual purpose package, for teaching and practical ‘field’ use.

For educational purposes, it assists animal health workers in understanding some elementary epidemiological procedures. INTRODUCTION and HELP pages are added to each program. A workbook with examples, suitable datasets, questions and answers is also available. EPISCOPE can also be used in lectures, when the VDU is replaced by a liquid crystal display in combination with an overhead projector.

For practical purposes, in addition to its value in designing field studies and analysing data, EPISCOPE could be used in simulation. It is especially useful to those who have no access to mainframe computers running suitable statistical packages, eg, practitioners. For example, EPISCOPE gives the same output for stratified case-control studies as a SAS procedure.

The data-analysis does not include multivariate techniques, such as analysis of variance or logistic regression. These techniques should not be used by the novice because the interpretation and the validity of the results requires deep insight into statistics (Rothman 1986).

EPISCOPE performs epidemiological calculations at great speed and it can evaluate alternatives by changing the input parameters. These characteristics might facilitate the widespread adoption of epidemiological procedures and stimulate motivation (Clarkson 1987).

A disadvantage of EPISCOPE is that the input in some programs consists of summary data, eg, frequencies in case-control and cohort studies. EPISCOPE does not calculate these summary data from the raw data. Summary data should be derived in another way, for example by hand or from a database program. This disadvantage is related to the fact that SUPERCALC is designed as a 'super calculator' and not as an efficient and easy-to-learn data handling program. Production of summary data using SUPERCALC involves a considerable knowledge of the SUPERCALC program.

Acknowledgements—The authors express their gratitude to Dr M. V. Thrufield of the University of Edinburgh, Department of Veterinary Clinical Studies, for his valuable assistance and critical comments.

References

BVA guides binder
AN attractive green binder is available for all BVA guides. Price £3.00 including postage from TGS Subscriber Services, 6 Bourne Enterprise Centre, Wrotham Road, Borough Green, Kent TN15 8DG, telephone 0732 884023, fax 0732 884034. Cash with order, please