Disease control analyses - stochastic modelling made easy

Managers of disease control programmes are commonly faced with analysing variable and/or imprecise data. Ignoring such issues might lead to erroneous decisions being taken, but handling such data presents some problems. Computer simulation using a model which incorporates an appropriate range of parameter values (i.e. a stochastic model) is an effective way of resolving these issues. Until recently such techniques have been the province of specialists. However, stochastic modelling software for personal computers has become available. These systems are easy to operate and interpret. This article describes how one such product has been used in analyses of the New Zealand tuberculosis control programme.

Variable and imprecise data are central problems of both science and commerce. Good examples of these issues are seen in the discipline of risk management. In the commercial arena efforts are directed at quantifying the probabilities that business objectives (usually profit) will or will not be met.[1] Likewise in the veterinary world, quantitative risk assessment has been increasingly used to calculate the probability of introducing animal disease by way of imported animals and animal products.[2]

How can such data be handled? Mathematical methods for calculation of variances of sums, differences, products and ratios are given in most standard statistical reference books. Techniques for handling very complicated data functions are also available. However, for those without an extensive mathematical background these methods are difficult to understand. Also, as the scope of the problem increases it may be unrealistic to attempt a purely mathematical solution.

Stochastic models

Computer models are used to approximate or mimic real-world systems, and models that include chance variation (i.e. stochastic models) are considered more biologically realistic than those that do not (i.e. deterministic models).[3] Stochastic models contain random number generators, and Monte Carlo sampling is used to allow for chance effects. Using such a system the investigator can include any observed variation or a possible range of effects.

Many stochastic models have been developed,[4] but because of their complexity they have generally been the province of research workers, not operational groups. However, personal computer modelling software which is easy to operate has become available. One such product (@RISK, Palisade Corporation, Newfield, New York) has been used extensively by disease control epidemiologists in New Zealand.

@RISK

The @RISK software package is a spreadsheet add-in. After starting the spreadsheet (Lotus 1-2-3, Lotus Corporation, or Excel, Microsoft Corporation) the @RISK program is automatically loaded. The software allows key variables to be entered in the form of specific distributions rather than as fixed values (e.g. mean, median, mode etc). The steps in carrying out the analysis are;

- Set up a model with inputs expressed as appropriate data distributions
- Execute repeated calculations, with a Monte Carlo sampling of any @RISK data distributions
- View the results of simulations (results, statistics and graphs).

MacDiarmid[2] has used this software in risk analyses on the importation of animals and animal products. This paper describes two examples of its application to routine analyses of disease control plans and their progress.

An analysis of the efficiency of diagnosing *Mycobacterium bovis* lesions by histopathology

In this study, lesions suspected of being due to *M.bovis* were collected from both cattle and deer at slaughter and submitted for histopathological examination and culture. Animal, herd and farm data were also retrieved from the National Disease Control Database.

One objective was to measure the efficiency of haematoxylin and eosin (HE) and Ziehl-Nielsen (ZN) histopathological examination. The sensitivities (SE) and specificities (SP) of the tests were calculated using culture as the gold-standard.[5] Multivariate logistic regression was used to investigate the association between SE and SP and the animal, herd, farm and laboratory variables.

Highly significant statistical effects were identified, and from these results a possible range of SEs and SPs for each test in deer and in cattle were estimated. The effect of these estimates of SE and SP on the association between positive predictive value (PPV) and disease prevalence (PREV) was modelled using the following equation[5];

PPV= (PREV * SE) / ((PREV * SE) + (1 - PREV) * (1 - SE))

Sensitivity and specificity were incorporated in the spreadsheet with @RISK as uniform distributions(i.e. minimum and maximum). A moderate negative correlation between SE and SP was also included. The mean PPV and the 95% confidence limits were calculated.

In the case of ZN histopathological examination in deer there were marked age, sex and laboratory effects. The possible ranges of sensitivity and specificities were wide. As would be expected, the resultant association between PPV and prevalence varied considerably (Figure 1).
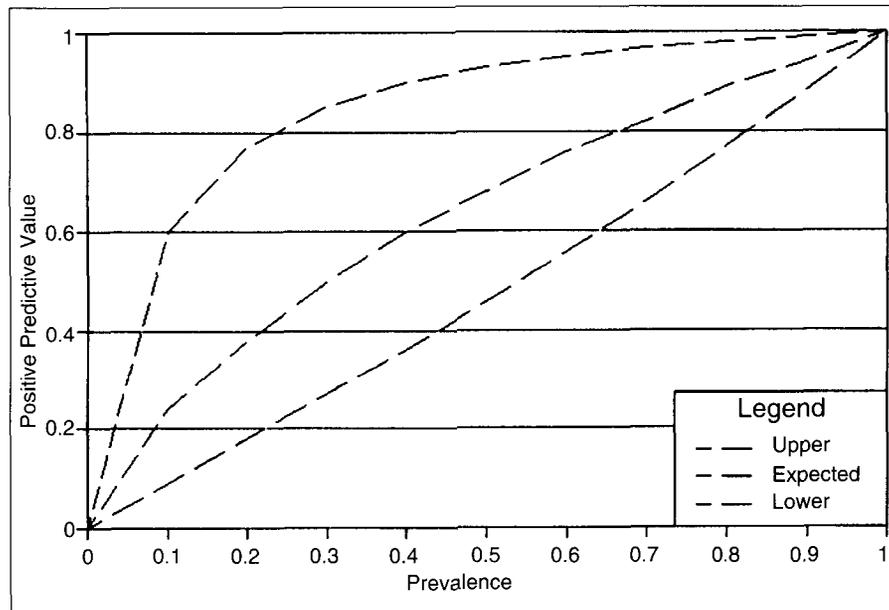
Development of a model of the tuberculosis eradication campaign

In New Zealand the campaign to eradicate tuberculosis from cattle and deer has been impeded by wildlife reservoirs of *M. bovis*. The most important vector is the brushtail possum (*Trichosurus vulpecula*), which was introduced from Australia early this century. However, the epidemiology of tuberculosis is complex, with other wildlife species implicated, and infection is also spread by livestock movement.

There has been considerable public debate over strategies to control tuberculosis. Possible measures include increased testing of cattle and deer, developing new field tests for tuberculosis, restricting livestock movements between farms, wildlife control and more stringent management of infected herds.

The @RISK software was used to model the effects of various policy options on the existing tuberculosis situation.

A path diagram which incorporated the main factors associated with the introduction and eradication of tuberculosis from herds in a district was developed and the final model contained several elements (Figure 2);

- Herd breakdowns caused by wildlife tuberculosis and the movement of infected livestock into the district
- Additional breakdowns resulting from spread within the district (i.e. spread from both new infection and existing infected herds)
- Total infected herds (i.e. sum of all breakdowns and residual infected herds)
- The number of infected herds tuberculin tested (calculated using the herd test frequency)
- The number of breakdowns detected and infected herds missed (estimated from the 'herd-test' sensitivity)
- Total movement control herds (calculated from the time on movement control).

The model was entered onto a spreadsheet. A data entry table was incorporated to allow for each of the input variables to be set annually (i.e. new tuberculosis infection resulting from wildlife and infected livestock entering the district, probability of spread within the district, probability of herd testing in a 6-month period, herd-test sensitivity, effect of abattoir surveillance, years on movement control and herds permanently on movement control).

The calculations were set up to run through one cycle each month, for a total of 20 years. At the end of each year, output parameters were copied to a summary table (i.e. probability of detecting an infected herd over a 6 month period, new infection caused by spread within the district, total infected herds, breakdowns expected, movement control number and prevalence).

Initially mean values were used as inputs. Later the spreadsheet was converted to operate with @RISK. In most cases the exact data distributions were not known and thus the simple triangular distribution was used (i.e. minimum, most likely, maximum).

The use of the @RISK model indicated that the results of any tuberculosis control strategy could vary considerably during the first few years of implementation.[6]

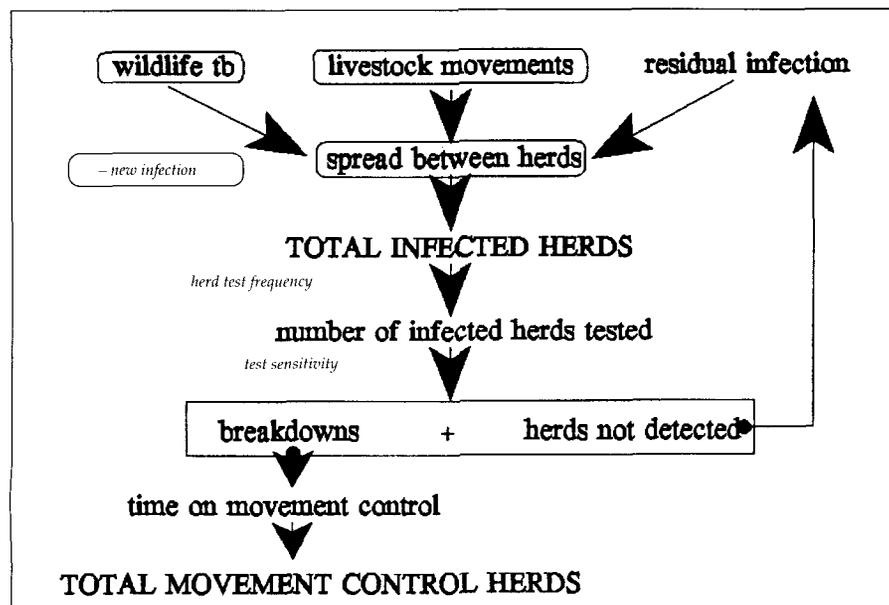*Figure 1:* ZN histopathological examination of suspect tuberculosis lesions from slaughtered deer. Positive Predictive Value versus Prevalence. Expected mean and 95% confidence limits.

Figure 2: Path model of tuberculosis infection and control.

Disease control analyses - stochastic modelling made easy

Conclusion

These two examples describe the process of developing stochastic models with the @RISK software and the valuable information that can be obtained from such models. This work also focuses attention of the requirement to understand the nature of data variability.

References

(1)   Anonymous. @RISK Risk Analysis and Simulation Add-In for Microsoft Excel or Lotus 1-2-3. Palisade Corporation, Newfield, New York.

(2)   MacDiarmid SC. Risk analysis and the importation of animals and animal products. Revue Scientifique et Tehcnique de l'Office International des Epizooties 12, 1093-1107, 1993.

(3)   Martin SW, Meek AH, Willeberg P. Veterinary Epidemiology. Principles and Methods. Pp. 195-216. Iowa State University Press, Ames, 1987.

(4)   Thrusfield M. Veterinary Epidemiology. Pp. 187-198. Butterworth & Co., London, 1986.

(5)   Dawson-Saunders B, Trapp RG. Basic & Clinical Biostatistics. Pp. 232-247. Prentice-Hall International Inc., New Jersey, 1994.

(6)   Ryan TJ. A Spreadsheet Model of Tuberculosis Control. Proceedings of the 7th International Symposium on Veterinary Epidemiology and Economics, Kenya, 1994 (In press).

T J Ryan
MAF Quality Management
Ruakura Agriculture Centre
Hamilton