The Kenya Veterinarian 18(2) (1994) 198-201

ESTIMATING SAMPLE SIZES FOR A TWO STAGE SAMPLING SURVEY OF THE LEVEL OF PROTECTION AGAINST FOOT AND MOUTH DISEASE IN NORTHERN THAILAND

CAMERON, A.R.[a], BALDOCK, F.C.[b], CHAMNANPOOD, P.[c], CLELAND, P.C.[c], JOHNSON, M.J.[d]

The Thai Government has recently greatly expanded its efforts to combat foot and mouth disease (FMD) through the use of vaccination. Recent studies in the north of Thailand (Cleland et al., 1993) have shown that high vaccination rates are required to establish and maintain a satisfactory level of herd immunity. It is intended to develop active surveillance methods to quickly assess overall levels of herd immunity as a measure of the effectiveness of the vaccination program. A two stage sampling design was chosen for the active surveillance of village cattle and buffalo. The two stage random sampling technique is a useful way to overcome survey design limitations due to the lack of a complete sampling frame and cost constraints. The estimation of the sample size required for such a design is complex. When accurate data relating to the target population are available, the required sample size may be estimated by repeated simulation of sampling experiments. The aim of the study described here was to estimate the number of villages and the number of cattle per village required to produce overall estimates of herd immunity at a prescribed accuracy for a target area.

MATERIALS AND METHODS

Empirical approach

The dataset: Existing data were used in a computer simulation study of sample sizes. Gleeson et al. (1993) studied the antibody response of cattle and buffaloes to vaccination with the serum neutralisation test in 21 villages in northern Thailand. At the commencement of the study, all available cattle in these villages (a total of 690) were tested for FMD titres against three serotypes (O, A, and Asia 1). Although limited in thier representative nature, these data were used to estimate the population parameters that formed the basis of the simulation exercise. Titres were assumed to be protective if the logarithm of the inverse of the titre was greater than 1.5. The mean number of animals per village was 32.86 with a standard deviation of 7.21. The proportions of the population with protective titres against strains A, O and Asia 1 were 11%, 44% and 24% respectively. At the village level, the mean of the proportion of protected animals in each village was 11%, 43% and 23%, with standard deviations of 9%, 16% and 22% respectively.

A population of cattle was simulated, and then repeatedly sampled to determine the accuracy of different sampling strategies. The simulated population was based on the observed level of protection against FMD serotype O because the proportion of the population protected was closest to 0.5, the value which yields the largest required sample size for a given accuracy.

Simulation of data: The distribution of village cattle populations and of the proportion of animals within villages with protective titres were assumed to be normal. The cattle populations of 100 simulated villages were randomly selected using a normal distribution, with a mean of 65.7 and standard deviation 14.4 animals. These values are double those found by Gleeson et al.(1993) to give populations large enough to use large

[a, c]Northern Region Veterinary Research and Diagnostic Laboratory, Hang Chat, Lampang, Thailand.
[b]Queensland Department of Primary Industries, Animal Research Institute, Locked Mail Bag No. 4, Moorooka, Queensland, 4150, Australia.
[c]South Australian Department of Agriculture, PO Box 1783, Port Lincoln, South Australia, 5606, Australia.
[d]Department of Mathematics, Macquarie University, North Ryde, NSW, Australia.

sample sizes without altering the distribution. Similarly, a value representing the proportion of protected animals was generated for each village, using a normal distribution, with a mean of 43% and standard deviation of 16%, as found in the original dataset. A Pascal program was then used to generate the individual cattle, being either protected or unprotected, using a simple random number generator. A random number that was less than or equal to the village proportion resulted in a protected animal, and a number more than the village proportion resulted in an unprotected animal. The resultant simulated data set contained 100 villages and 6574 cattle. The mean village size was 65.7 cattle with a standard deviation of 13.7 and a range from 32 to 104. The mean of the village proportions was 41.8% with a standard deviation of 16.6%. There were 2699 protected animals in the dataset, giving an overall proportion of protected animals (P) of 41.1%.

Sampling technique: Sampling was simulated using a program written in Pascal. The entire dataset was analysed, and the true population proportion of protected animals (P) calculated. Then, for every combination of n (number of villages) and m (number of animals per village), the following sampling procedure was used. A random number generator was used to generate n random numbers between 1 and N (the total number of animals in the population), to select villages without replacement and with probability proportional to size. For each selected village, m random numbers between 1 and M (size of village) were generated to select animals without replacement. For each village, the estimated number of protected animals within the village was calculated by multiplying the proportion of protected animals in the sample by the size of the village. The estimated population proportion (p) weighted proportional to village size was calculated by summing the estimated number of protected animals for each village, and dividing by the total number of animals in the sampled villages.

If the result p was greater than P-0.1 and less than P+0.1, then the result for the trial was said to be correct to within ± 10%, and similarly for ± 5% (i.e. P-0.05 and +0.05). Each trial was repeated 1000 times, and the cumulative numbers of correct results at the 10% and 5% levels were recorded.

Analytic Approach

The relationship between the number of villages sampled, the number of animals in each village sampled, and the accuracy of the estimate is dependant on the variation in level of protection between villages and the variation within villages. This may be described as V (total variance) which is equal to the variance among villages plus the variance within villages , or:

$$V=\frac{\sigma^2}{n}+\frac{p(1-p)}{m}$$

where p=proportion of animals protected, n=number of villages, m=number of animals per village, and σ^2= variance of the village means. The confidence interval (CI) for an estimate p is equal to:

$$CI=Z_{\frac{\alpha}{2}}\sqrt{\frac{V}{n}}$$

Combining formulae 1 and 2, and expressing in terms of m, this becomes:

$$m=\frac{p\,(1-p)}{n\left(\left(\frac{CI}{Z_{\frac{\alpha}{2}}}\right)^2-\frac{\sigma_p^2}{n^2}\right)}$$

In the situation described in the simulation experiment, the confidence interval is 0.1 (± 10% of the true proportion), Z = 1.96 (estimate is correct 95% of the time), p = 0.44, and $\Gamma^2=0.0256$. The optimum combination of m and n may be calculated, if the costs associated with sampling are known, by using the formulae of Cochran (1977).

RESULTS

The proportion of correct results at the 10% level for various combinations of m and n, yielded by the empirical approach, is shown in Fig. 1. Figure 2 shows all the combinations of m and n that yield various levels of accuracy, as determined by the analytic approach.

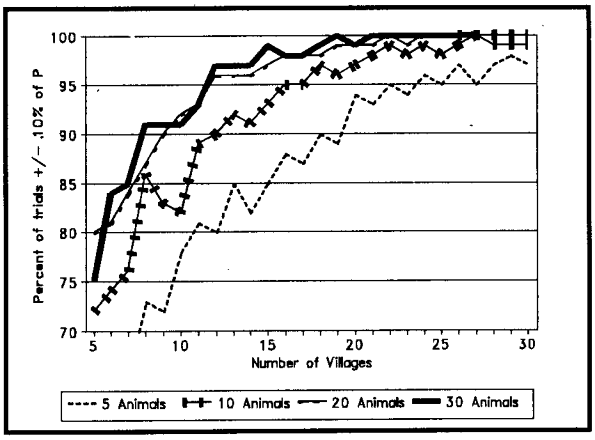

Fig.1 . The proportion of trials with different numbers of villages and animals per village within ± 10% of P.

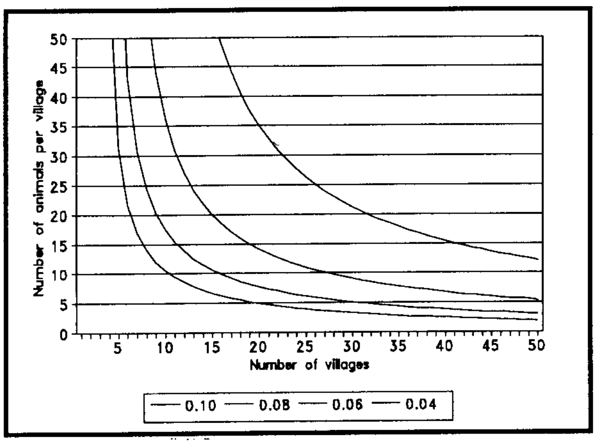

Fig.2. The number of villages and animals per village required for results of given levels of accuracy.

DISCUSSION

The two approaches used to estimate the required sample sizes for the projected survey yield similar results. However, estimates yielded by the analytic approach suggest somewhat lower numbers of either animals or villages are required to provide the same accuracy. For instance, when sampling 20 villages, the analytic approach suggests that only around 5 animals are needed to be 95% confident of obtaining results within 10% of the true proportion. The empirical sampling exercise suggests that this combination would yield a result within 10% of P only 93% of the time. As the number of villages sampled decreases, the discrepancy between the two techniques increases. The assumption of normality in the distributions of the size of each village and the proportion of protected animals between villages is not completely justified. The Wilk-Schapiro statistic for these variables is 0.882 and 0.957 respectively, indicating that the distribution of the size of the villages is non-normal. However, the results of the calculations in this exercise are not very sensitive to this departure from normality and it can be safely ignored.

The results of both approaches would tend to support the use of the rapid epidemiological assessment of two stage sample sizes as described by Henderson and Sundaresan (1982). The results presented here suggest that the scheme of 30 clusters, with seven sampling units in each, would be very likely to produce results within the 10% confidence interval desired in this situation.

REFERENCES

Cleland P.C., Baldock F.C., Gleeson L.J., Chamnanpood P., 1993. A modelling approach to the investigation of vaccination strategies for foot and mouth disease. In: Copland J.W., Gleeson L.J. and Chamnanpood P. (Editors). Diagnosis and epidemiology of foot and mouth disease in South-East Asia. Proc. International Workshop 6-9 Sept 1993, Lampang, Thailand. ACIAR, Canberra, Australia.

Cochran, W.G., 1977. Sampling Techniques (3rd edition). John Wiley and Sons, New York. 428pp.

Gleeson L.J., Robertson M.D., Doughty W.J., Chamnanpood C., Cleland P.C., Chamnanpood P. and Baldock F.C., 1994. Serological responses of village cattle and buffalo in northern Thailand to a newly introduced trivalent foot and mouth disease vaccine. In: Copland J.W., Gleeson L.J. and Chamnanpood P. (Editors). Diagnosis and epidemiology of foot and mouth disease in South-East Asia. Proc. International Workshop 6-9 Sept 1993, Lampang, Thailand. ACIAR, Canberra, Australia.

Henderson, R.H. and Sundaresan T., 1982. Cluster sampling to assess immunization coverage - a review of experience with a simplified sampling method. Bull. WHO, 60: 253-260.