Epidémiol. santé anim., 1997, 31-32

RAPID, INEXPENSIVE DISEASE OCCURRENCE ESTIMATES IN DEVELOPING COUNTRIES : THE ROLE OF SURVIVAL ANALYSIS AND CAPTURE-RECAPTURE TECHNIQUES

Cameron A.R.[1], Chamnanpood P.[2], Khounsy S.[3]

Les données d'incidence des maladies sont traditionnellement basées sur les relevés exhaustifs des cas. Dans beaucoup de pays en voie de développement cependant, les systèmes d'enregistrement des morbidités sont inadaptés et sous-estiment significativement le nombre total de cas. Deux approches alternatives pour une estimation rapide, non-biaisée et peu coûteuse de l'occurrence des maladies sont présentées ici.
La première approche est appropriée aux maladies survenant sous la forme d'épizootie cyclique. Un questionnaire rétrospectif est utilisé à l'échelle du troupeau ou du village, et les données sont traitées en utilisant l'analyse de survie afin d'estimer les courbes de survie. L'utilisation de techniques adéquates d'entretien des éleveurs villageois assure que l'information rétrospective collectée est aussi précise que possible. On ne peut pas en tirer des mesures classiques des incidences pathologiques, mais les méthodes de comparaison des courbes et de quantification des différences sont présentées. Des exemples d'enquêtes utilisant les techniques de comparaison des différences dans les occurrences pathologiques en Thaïlande et au Laos sont présentés.
La seconde approche utilise une combinaison de données acquises passivement telles qu'elles sont rassemblées dans des registres de morbidité, et des données d'enquête. Ces deux comptages différents des cas de maladie sont analysés par des méthodes de capture-recapture afin de procurer une estimation ajustée du nombre total de cas. Ces estimations peuvent être utilisées pour calculer directement les taux d'incidence des maladies.

INTRODUCTION

Measures of disease occurrence are required by veterinary authorities for a variety of reasons, such as priority setting, control program planning and international reporting. Passive disease reporting systems are the main source of this information in many developing countries. These systems suffer from two important problems. First, only data on diseased animals is available, with no reference to the population at risk. This means that epidemiological measures such as prevalence or incidence rates cannot be calculated. Second, the total number of occurrences of disease is usually underestimated, due to under-reporting (Ogundipe et al., 1989). Incidence rates based on this data may be severely biassed and reliable inferences about the population are not possible. The solution to this problem is the use of active surveillance, the use of special purpose targeted surveys to gather population-based information of quantifiable accuracy. Due to practical and cost considerations, large scale surveys of this nature are usually cross-sectional, and are therefore only able to produce estimates of point prevalence, not disease incidence rates.

The lack of reliable disease incidence data in developing countries may severely limit the veterinary authorities' understanding of the disease situation. Consider, for example, an active surveillance program using sero-surveys to monitor the effectiveness of a national disease eradication campaign based on vaccination. For many diseases (such as Foot and Mouth Disease), it is difficult or impossible to distinguish between antibody titres derived from vaccination, and those derived from natural exposure to the pathogen. Survey results indicating that a high proportion of the livestock population have high antibody titres may indicate that the vaccination program has been very successful. Alternatively, it may indicate that the program has been a complete failure and the rate of natural infection is extremely high. Without reliable measures of disease incidence rates, it may be very difficult to reliably distinguish between these two situations. To be sure that a vaccination programme is working, there must be an increase in the seroprevalence and a decrease in the disease incidence rate.

Traditionally, reliable estimates of disease incidence rates require accurate and complete records from a relatively large population over an extended period. Collecting and maintaining such information is not currently feasible in many developing countries. In addition, the time required means that estimates are slow to gather and do not necessarily reflect the current situation.

This paper discusses two approaches to estimating disease incidence rate using rapid, inexpensive data collection techniques. The first uses retrospective village outbreak questionnaires and event history analysis (analogous to survival analysis, but working backwards in time). The second uses two sources of data on disease events, and analyses them with capture/recapture methodology.

1 Lao-Australian Animal Health Project, PO Box 7042, Vientiane, Lao PDR
2 Northern Veterinary Research and Diagnostic Centre, Hang Chat, Lampang 52190, Thailand
3 Animal Health Division, Department of Livestock and Fisheries, Ministry of Agriculture and Forestry, Vientiane, Lao PDR

RETROSPECTIVE VILLAGE OUTBREAK ANALYSIS

This technique differs from traditional incidence estimation in several key ways. First, it does not measure disease occurrence in individual animals, but rather considers the herd or village as the unit of interest. The disease event analysed is a village outbreak. Second, it is limited in its applicability to disease which can occur repeatedly, and in outbreak form. The diseases must be significant enough to be easily recalled, and be able to be reliably diagnosed by villagers. Third, it does not produce a traditional incidence estimate, but a curve describing the outbreak experience of the population. This curve, however, can be analysed and compared in the same way as incidence estimates.

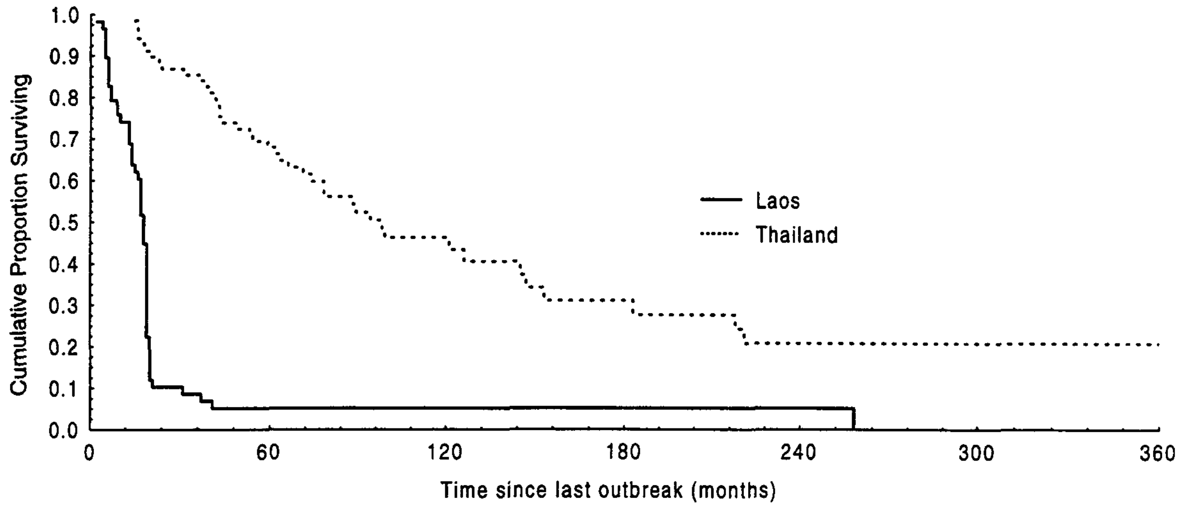
The technique involves the use of a simple retrospective village questionnaire survey. During a group interview of village livestock owners, villagers are asked to recall, as accurately as possible the time of the most recent outbreak of the disease in the village. If no outbreak can be recalled, the earliest date since which villages are sure there has been no outbreak is recorded.

Collection of outbreak occurrence data based on the memory of villagers, and potentially stretching back over many years, presents some problems and naturally introduces some inaccuracies. Group interview techniques are available to aid in the accuracy of recall, such as the development of village histories and calendars to aid with the accurate timing of an historical event, and the participation of many livestock owners so that the "group memory" of the village can be tapped. The use of this technique with FMD is particularly appropriate, as the disease has a dramatic and memorable clinical manifestation, and occurs in distinct outbreak form in a cyclic pattern.

Survival analysis techniques are used to analyse the results of such a survey. The disease experience of the villages surveyed can be summarised in the form of a Kaplan-Meier survival curve (Lee, 1992). This data cannot be reliably converted into the more familiar incidence rate, but reflects the same information. It can, however, be used to detect changes in disease occurrence over time, or differences between two areas. The log-rank test (Tibshirani, 1982; Lee, 1992) is a statistical test which compares two survival curves to determine if they are likely to have been drawn from the same population. Differences between curves can be quantified using the hazard ratio, which is analogous to the risk ratio (Armitage and Berry, 1994). The hazard ratio may be used to quantify progress in a disease eradication campaign, by comparing data from an area with that from a previous year. Problems in regional coverage of a campaign may be identified by comparing data from different areas.

During the development of this technique, a survey of 69 villages was conducted in December 1995 in three provinces of northern Thailand, and a second survey of 35 villages in November 1996 in seven districts of Vientiane Prefecture, in Lao PDR. The two Kaplan-Meier survival curves produced are shown in Figure 1.

Figure 1
Kaplan-Meier survival curves from surveys of village FMD outbreaks in Thailand and Laos



The two curves are clearly different, and analysis with the log-rank test produces an L- statistic of 6.97 ($P<0.001$). This indicates that it is extremely unlikely that the two samples came from populations with the same experience of FMD outbreaks. This difference is quantified by a hazard ratio of 4.16 (95% confidence interval 2.8 - 6.2), indicating (loosely speaking) that villages in Laos have approximately four times the risk of an outbreak of FMD than those in Thailand. Thailand has been conducting an extensive, nationwide vaccination campaign against FMD for several years, whereas in neighbouring Laos, virtually no measures are in place to control the disease.

The weaknesses in this technique that may be criticised by purists are, in fact, its strengths and features that make it appropriate for use in developing parts of the world. The data is gathered by group interviews and depends on people's memories, opening the way to inaccuracies. However, this also makes the survey very rapid and inexpensive to conduct, requiring no laboratory support, and able to produce usable measures of disease occurrence in a period of weeks. When no other data exists, or existing data sources (such as passive outbreak reports) are severely biassed, and funds are limited, rapid inexpensive surveys that provide quantitative measures of disease occurrence are badly needed. As long as the potential inaccuracies and limitations of this approach are recognised and minimised through appropriate use and careful interview technique, the use of retrospective outbreak questionnaires can fill an important gap in our disease knowledge.

CAPTURE/RECAPTURE METHODOLOGY

This technique, developed mainly for wildlife population estimates, can be applied to the problem of estimating disease incidence in domestic animals. If one needs to estimate the total number of fish in a lake, it is possible to use a combination of the results of two surveys. Firstly, a number of fish are caught, tagged and released. Then a second group of fish are caught. The total number of fish caught in the first and second rounds, and the number of fish caught in both (ie tagged fish caught in the second round) can be used to estimate the total number of fish in the lake. In the same way, the results of two independent disease surveys can be combined to yield an estimate of the total number of disease events in a given period.

Passive disease reporting systems provide the first source of information. To be useful, these records must clearly identify the village or animal in which the disease event occurred, and the time at which it occurred. Although this approach may be used for both individual animal and village outbreak incidence rate estimation, problems in individual animal identification limit its use at this level. The second source of information is derived from a special-purpose survey, perhaps in combination with a seroprevalence survey. Using group interview techniques described above, details of previous disease events that have occurred in the village over a fixed period (usually one or two years) are recorded. These records are compared to those of the passive reporting system, and disease events which appear on both lists are identified and counted. An estimate of the total number of outbreaks in the designated time is given by:

$$N = \frac{n_1 n_2}{n_{12}}$$

where N is the total number of outbreaks over the period, n_1 is the total number of outbreaks identified in the first source (passive reports), n_2 is the number from the second source (survey results) and n_{12} is the number of matching outbreaks appearing in both sources. If the total number of villages in the area is known, the estimated total can be used to calculate a crude incidence rate in the form *outbreaks per hundred villages per year.*

There are many limitations to this approach. Unless the number of outbreaks appearing on both lists, n_{12}, is relatively high, the estimate may be very imprecise (see Seber (1970) for confidence interval calculation). For stable estimates, both the number of reported outbreaks and the number of outbreaks identified through the survey should be large. In addition, the estimate may be biassed due to the way in which outbreaks are reported. This bias will usually result in an underestimate of the total number of outbreaks, and may be able to be controlled for by more sophisticated techniques (Sekar and Deming, 1949) or the use of more than two lists (Yip et al., 1995). Despite these drawbacks, the estimate of disease incidence produced will be far more accurate than that derived from passive reports.

DISCUSSION

The two methods describe provide an opportunity to collect simple, readily available information rapidly, which can be used to generate epidemiologically sound estimates of disease incidence rates or analogous measures. The main advantage of these techniques is their low-cost and simplicity. If active serosurveillance is being used to monitor the health of livestock, village group interviews can be conducted, at virtually no extra cost, to gather the required information. The techniques offer a means to improve the understanding and control of important livestock diseases in developing countries.

ACKNOWLEDGEMENTS

This research was funded by the Australian Centre for International Agricultural Research and conducted in collaboration with the Thai Department of Livestock Development and the Lao Department of Livestock and Fisheries. Angus Cameron is supported by a Junior Research Fellowship from the Australian Meat Research Corporation.

BIBLIOGRAPHY

Armitage P., Berry G. 1994. Statistical Methods in Medical Research, 3rd ed. Blackwell Science, Oxford.

Lee E.T. 1992. Statistical methods for survival data analysis, 2nd ed. John Wiley & Sons, Inc, New York.

Ogundipe G.A.T., Oluokun S.B., Esuruoso G.O., 1989. The development and efficency of the animal health information system in Nigeria. Preventive Veterinary Medicine 7, 121-135.

Seber G.A.F., 1970. The effect of traph response on tag recapture estimates. Biometrics 26, 13-22.

Sekar C.C., Deming W.E., 1949. On a method of estimating birth and death rates and the extent of registration. Journal of the American Statistical Association, 101-115.

Tibshirani R., 1982. A plain man's guide to the proportional hazards model. Clinical and Investigative Medicine 5, 63-68.

Yip P.S.F., Bruno G., Tajima N., Seber G.A.F., Buckland S.T., Cormack R.M., Unwin N., Chang Y.-F., Fienberg S.E., Junker B.W., LaPorte R.E., Libman I.M., McCarty D.J., 1995. Capture-recapture and multiple record systems estimation II: Applications in human diseases. American Journal of Epidemiology 142, 1059-1068.