The National Animal Health Monitoring System in Michigan. II. Methodological Issues in the Estimation of Frequencies of Disease in a Prospective Study of Multiple Dynamic Populations

H. SCOTT HURD and JOHN B. KANEENE

The Division of Epidemiology, Department of Large Animal Clinical Sciences, College of Veterinary Medicine, Michigan State University, East Lansing, MI 48824-1316 (U.S.A.)

ABSTRACT


Procedures are proposed for the computation of disease frequency measures and of their associated variances from data collected through prospective study of multiple dynamic cohorts (herds) with a National Animal Health Monitoring System. Estimates of either the annual incidence density for a group of herds or the 1-month risk of disease can be calculated from the same within-herd measure of monthly incidence density.

It is proposed that the choice of measure to be estimated depends on the intended use of the information. Risk estimates are appropriate for producers and clinical health professionals making decisions at the animal or herd level. Incidence density measures are appropriate for extrapolation to reference populations used for state- and regional-level decision-making.

INTRODUCTION

Disease monitoring activities in livestock populations can have a variety of forms, with a variety of goals and users (Beal, 1983). Some activities observe prevalent cases of disease at slaughter (Willeberg, 1978; Lloyd and Schwab, 1987) or upon submission to a diagnostic laboratory (Davies, 1978). Other activities might observe disease occurrences as a result of continuing herd management and disease control programs (Stephens et al., 1982; Bartlett et al., 1986; Dohoo and Stahlbaum, 1986). The shortcomings of these activities for gaining estimates of disease frequency that are statistically valid (and hence useful for extrapolation to a source population) have been discussed (Beal, 1983). Efforts to overcome these shortcomings have resulted in the antecedent
Minnesota Food Animal Disease Reporting System (Diesch, 1982, 1983) and the present National Animal Health Monitoring System (NAHMS). These projects are unique in that herds are purposefully selected for follow-up. In these situations, multiple cohorts (herds) are observed repeatedly in a prospective longitudinal study, to obtain an estimate of disease frequency that can be extrapolated (with defined confidence limits) to a reference population at the state, regional, or national level.

By comparison, most follow-up studies in human epidemiology involve observations on only one or two cohorts of individuals. The purpose of these studies is usually to test a hypothesis on the effect of an exposure or risk factor on the occurrence of disease in an individual, not the estimation of a population parameter (Kleinbaum et al., 1982; Susser, 1985). Consequently, most statistical procedures have been developed around estimation and variance calculations of effect measures such as the risk ratio, odds ratio, and risk difference (Fliess et al., 1976; Rothman, 1986). In animal populations some work has been done to estimate the prevalence of disease (Beal, 1985; Farver et al., 1985); however, this may not be directly applicable to incidence data from follow-up studies (Chiang, 1961).

Estimation procedures need to be documented that can be applied consistently to estimate the frequency of disease in a population which consists of multiple dynamic subpopulations (herds or cohorts). As the NAHMS is a national effort in the U.S.A., it is important that these and other methodologies should be standardized, for comparability between states. The objectives of this paper are to: (1) raise the issues involved in disease frequency estimation; (2) propose a criterion for the determination of the appropriate disease frequency measure to be estimated; (3) propose methods for the computation of sample estimates and variances of these frequency measures.

METHODS

Herd selection and data collection

Sixty \((n = 60)\) of the 6012 dairy cattle herds in Michigan were selected randomly for 1 year of follow-up. A detailed account of the selection of these herds, data collection tools, and the data collected are presented in the first paper of this series (Kaneene and Hurd, 1990). Using a multistage sampling procedure, a stratified sample of 60 herds \((i)\), with probability proportional to size, was selected. The strata \((j)\) were: stratum 1, 10–49 adult cows (milking and dry); stratum 2, 50–99 cows; stratum 3, 100–199 cows; stratum 4; \(\geq 200\) cows. The herd was the sampling unit, and all animals within the herd were observed for the occurrence of disease. Each farm was visited at least monthly for 1 year to collect data on the frequency and cost components of dairy cattle diseases. The data collected at each visit included a count of the number of cases of disease
that occurred during the month. Recurrence of disease in a previously recovered animal was counted as a new case of the disease. The number of animals in each age group at the end of the month was recorded, as were the numbers of animals added to and removed from the herd. Information on individual animals (such as calving date and age) was also collected in the Michigan project. However, as this information is not usually collected in the standard NAHMS design, it will not be considered in this paper.

End-product to be estimated

Two general types of disease frequency have been recognized: the risk rate and the true rate or incidence density (Kleinbaum et al., 1982; Bendixen, 1987; Martin et al., 1987). We propose that the choice of measure to be estimated depends on the intended use of the estimates – with constraints based on the type of data available. Incidence densities are useful for population estimates and macroeconomic decision-making by state- and national-level policy-makers. The risk rate is useful for the selection of appropriate treatments and for personal decisions on health-related behaviors – in other words, micro-level decision-making by producers and clinically concerned health professionals.

The incidence density is a meaningful measure of the experience of the population group (Chiang, 1961), and as such is more appropriate for extrapolation (Alderink, 1986). Leech (1971), when discussing disease monitoring systems in Britain, commented that incidence density measures are more meaningful than risk estimates. The animal disease surveillance program in Minnesota estimated disease frequency as incidence densities (Diesch, 1983), and incidence densities of mastitis have been reported by Bendixen et al. (1988). However, the incidence density has no application at the level of an individual animal, and is not very useful to a producer or clinical epidemiologist (Miettinen, 1976; Bendixen, 1987). (The exception is when incidence density measures are used to determine the relative risk of a given exposure.)

Morgenstern et al. (1980) defined risk as the conditional (a priori) probability of disease occurrence in an individual. Risk estimates of common dairy cattle diseases have been reported by many investigators (Martin et al., 1975; Simensen, 1982; Dohoo et al., 1983; Erb et al., 1984; Bartlett et al., 1986; Waltner-Toews et al., 1986; Bendixen, 1987; Curtis et al., 1988). Although some of these reports may use the term "incidence rate", examination of the calculations shows that the measure is actually a risk. These reports are generally aimed at the producer or the clinician and relate to the question of how individual factors may affect an animal’s risk of disease. These risk estimates might be useful for decisions on the implementation of some intervention, such as a vaccination program, which must include consideration of the expected probability (i.e. the risk) that an animal and/or herd will be infected.

Bendixen (1987) discussed how the type of data constraints the choice of
measure to be estimated. The risk (or cumulative incidence) is appropriate for fixed populations and for some dynamic populations if the risk period is well defined; the incidence density is appropriate for dynamic populations when disease occurrence is not restricted to a specific time. The latter applies to the NAHMS.

**Herd calculation of disease frequency**

Calculation of the frequency of disease in a given herd for a given month can be implemented in a variety of ways, depending on whether the population at risk is assumed to be of fixed size during the observation interval or of changing size. If the population is fixed or animals are at risk for a given period of time, then the cumulative incidence (definable or proportion of animals affected) can be calculated (Elandt-Johnson, 1975; Kleinbaum et al., 1982). This number represents the risk of disease for a given time period. The risk is conditional on new animals not entering the population and on only one occurrence of the disease per animal being considered.

In a situation such as the NAHMS, the population size may change dramatically during the observation period, and animals can be affected more than once. The former was particularly true for calves in the Michigan study. In this situation, the incidence density is recommended (Miettinen, 1976; Kleinbaum et al., 1982; Rothman, 1986; Bendixen, 1987). Also, in cases where data on individual animals are not available, it is not possible to calculate measures such as the lactational incidence rate or periparturient risk rates, as it is not possible to determine which animals have calved recently and are at risk. The formula for the incidence density ($ID_{ijm}$) for one disease for 1 month for the $i$th herd in the $j$th stratum is shown in eqn. (1). This is a modified version of the actuarial method as shown by Elandt-Johnson (1977).

Incidence densities for some diseases may be biased downward because of the inclusion of animals in the denominator that are no longer at risk for that particular disease. Because of the nature of the data collected, all animals in the herd (in a given age group) contribute to calculation of the animal-months. Collection of individual information for all age groups would allow better estimation of frequencies and provide better data for research applications, such as identification of important individual risk factors.

$$ID_{ijm} = \frac{\text{no. of cases}}{\text{animal-months}}$$

$$ID_{ijm} = \frac{\text{no. of cases of disease (X) during current month}}{(\text{no. of animals at risk at end of previous month} - \frac{1}{2}\text{withdrawals} + \frac{1}{2}\text{additions})}$$

where
Withdrawals = no. sold + no. died due to other disease + no. transferred to different age group
Additions = no. purchased + no. transferred in from other age groups

Population estimates of annual incidence

As it has been proposed that the incidence density (true rate) is the most meaningful expression of disease frequency for macro-level decision-making, the calculation of a population estimate of this measure will be discussed. For population estimates of disease, it may be reasonable to aggregate the results of monthly observations into one annual figure for each herd. The monthly incidence densities ($ID_{ijm}$) are aggregated easily as shown in eqn. (2) (Rothman, 1986). This term will be referred to as the "annual" incidence density ($aID_{ij}$) — which is not as accurate as calling it the incidence density expressed in terms of animal-years, but is more convenient. This annual incidence density ($aID_{ij}$) represents the average force of morbidity observed over 1 year of repeated monthly observations. It assumes a constant rate of disease for the entire year; hence, seasonal fluctuations are ignored. This may present a problem, as some diseases did show noticeable seasonal fluctuations (Kaneene and Hurd, 1990).

\begin{align*}
aID_{ij} &= \text{"annual" incidence density for the } i\text{th herd in the } j\text{th stratum, expressed per 100 animal-years} \\
&= \left( \frac{\sum_{m=1}^{12} \text{cases}}{\sum_{m=1}^{12} \text{animal-months}} \times 12 \right) \times 100 \tag{2}
\end{align*}

If the annual incidence density ($aID_{ij}$) represents the unit of measure for one herd, then a summarization of the frequency for many herds is required to estimate the rate for a population. It is proposed that the data should be stratified according to herd size (Beal, 1985), as this is assumed to affect management and comparability of disease rates. For this reason, discussion of the summarization of rates will not proceed beyond the herd size stratum ($j$) level, although an overall frequency estimate is estimable if desired. The average stratum-specific annual incidence density ($aID_j$) is computed as in eqn. (3).

It is the weighted average of the herd-specific annual IDs ($aID_{ij}$). The weights are those used in standard sampling theory for estimation of proportions (Cochran, 1977; Levy and Lemeshow, 1980). Other weights (such as the inverse of the variance) could be considered, as the incidence density is not a binomial proportion (Kleinbaum et al., 1982; Rothman, 1986). It is also reasonable that a finite population correction factor $n_j/N_j$ (for the herd sampling) should be considered; however, this was dropped as the number of herds sam-
pled \( n_j \) was relatively small compared with the number of herds in the state \( N_j \). Dropping this factor also simplifies the equations for discussion.

\[
\bar{aID}_j = \text{estimate of the "annual" incidence density of the } j\text{th herd size stratum}
\]

\[
= \frac{\sum_{i=1}^{n} W_{ij}aID_{ij}}{n}
\]

where

\[
W_{ij} = \frac{aNAR_{ij}}{\sum_{i=1}^{n} aNAR_{ij}}
\]

\( n = \text{no. of herds in the } j\text{th stratum, and} \)

\[
aNAR_{ij} = \text{no. of animal years} = \frac{\sum \text{animal-months}}{12}
\]

**Variance estimates of annual incidence**

If the multistage sampling procedure is viewed as a cluster sample (in which herds are the clusters), then variation will be contributed from within herds and between herds (Beal, 1985; Alderink, 1986; Farver, 1987). However, it might be argued that, as we are observing all the animals within the herd for disease, there will be no within-herd variance in the \( aID_{ij} \). This is equivalent to saying that we have perfect knowledge of the rate of disease in that herd and that it is not subject to any random variation. Chiang (1961) suggests that we are considering a stochastic phenomenon (disease) which is subject to chance; as such, a within-herd variance should be considered.

As the incidence density \( (aID_{ij}) \) is neither a binomial proportion nor a probability function, its variance must be approximated by its relationship to the probability (risk) of disease in a given herd. The relationship of incidence density to risk for a given period has been described as shown in eqn. (4) (Morgenstern et al., 1980). Application of this functional relationship results in a variance estimator for \( aID_{ij} \) as shown in eqn. (5). Annual number at risk \( (aNAR) \) represents the average number of animals in the herd for the year. The variance of the average annual incidence density \( S^2(aID_j) \) for the \( j\text{th stratum} \) is a function of the weighted individual herd variances \( S^2(aID_{ij}) \) and the between-herd variances as shown in eqn. (6):

\[
\text{Risk} = 1 - \exp(-ID) \quad (4)
\]

\[
S^2(aID_{ij}) = \frac{aID_{ij}[1 - (1 - \exp(-aID_{ij}))]}{aNAR_{ij}} \quad (5)
\]
DISEASE FREQUENCY ESTIMATION IN MICHIGAN

\[ S^2(aID_j) = \frac{\sum_{i=1}^{n} (aID_{ij} - aID_j)^2}{n_j} + \sum_{i=1}^{n} W_{ij} (aID_{ij}) \]  

= sample estimate of the variance of the “annual” incidence density for ith herd in jth stratum

where

\( n_j = \text{no. of herds in jth stratum} \)

Risk estimation

An estimate of the probability of occurrence of a disease on a farm should aid producers in planning health care and other management changes. The risk estimate must be defined for a specific time period – for example, the 1-month risk of disease, the 1-year risk, or the lactational risk. For some diseases, the 1-month risk (\( R_{ijm} \)) will vary seasonally. As these risk estimates were reported to the participating NAHMS producers they were not aggregated over time. Therefore, equations shown are for stratum estimates of the 1-month risk (\( \bar{R}_{ijm} \)) and its variance.

The 1-month risk (\( R_{ijm} \)) for a given herd can be approximated from the incidence density for 1 month. If the \( ID_{ijm} \) is less than 0.10 and the time period is short, the risk and ID can be assumed to be equal (Kleinbaum et al., 1982). If these assumptions do not hold, the risk can be approximated using eqn. (4). In most cases in Michigan, these assumptions held true, but occasional outbreaks of diarrhea and respiratory disease in calves did result in large IDs (some greater than 1.0) – which is acceptable for the incidence density measure (Rothman, 1986). Therefore, the conversion in eqn. (4) is recommended as a routine procedure for the NAHMS. An estimate of the stratum 1-month risk (\( \bar{R}_{ijm} \)) and its variance \( [S^2(\bar{R}_{ijm})] \) can proceed as in eqns. (7)–(10) – treating the risk as a binomial proportion (Elandt-Johnson, 1977; Martin et al., 1987). Use of the cluster method for computing a variance should be considered (Beal, 1985), but one is then forced to assume that \( R_{ijm} \) represents an observed number of positive cases (\( a_{ijm} \)) out of an observed number of sampled cases.

\[ R_{ijm} = 1 - \exp (-ID_{ijm}) \]  

\[ R_{ijm} = 1 \text{-month risk of disease (X) for the ith herd in the jth stratum, for the } \]  

\[ m \text{th month} \]
\[ S^2(R_{ijm}) = \frac{R_{ijm}(1-R_{ijm})}{m_{ijm}-1} \]  

where

\[ m_{ijm} = \text{hypothetical no. of animals at risk in the } i \text{th herd in the } j \text{th stratum for the } m \text{th month} \]

\[ S^2(R_{ijm}) = \text{sample estimate of the variance of the risk from the } i \text{th herd in the } j \text{th stratum for the } m \text{th month} \]

\[ \bar{R}_{jm} = \frac{\sum_{i=1}^{n} W_{ijm} R_{ijm}}{\sum_{i=1}^{n} W_{ijm}} \]

where

\[ W_{ijm} = \frac{m_{ijm}}{\sum_{i=1}^{n} m_{ijm}} \]

\[ i = 1 \text{ to } n; \ n = \text{no. of herds in the } j \text{th stratum} \]

Further summarization of the stratum risk (\( \bar{R}_{jm} \)) estimates could be accomplished in two separate ways: computation of a mean 1-month risk for 12 months of observation (\( \bar{R}_j \)), and computation of an annual risk of disease (\( \text{annR} \)). The mean monthly risk for 1 year of observation could be computed as the sum of all the weighted mean risk estimates (\( \bar{R}_{jm} \)) divided by the number of months of observation. This will represent the average 1-month risk of disease; its value may be limited as this measure tends to ignore the fact that multiple cases of disease can occur in the same animal and that the occurrence of disease in an animal or risk (\( R_{ijm} \)) for a herd may not be independent from one month to the next. Computation of a variance for these risk estimates is not meaningful, as the risk is likely to vary significantly within the 12 months of observation as a result of seasonal effects, vs. random variation.

The annual risk of disease (\( \text{annR} \)), which could be a valuable measure, would be computed according to eqn. (11) (Kleinbaum et al., 1982). This measure is appropriate only if animals are at risk for the entire 12-month period. This
DISEASE FREQUENCY ESTIMATION IN MICHIGAN

Does not apply for calves which mature into a different age category, or for periparturient cow diseases. For diseases with a limited risk period, the annual risk (\(\text{annR}\)) will probably be inaccurate as the model assumes that animals are at risk for the full 12 months and the risks are derived from IDs which are biased downward for these diseases.

\[
\text{annR} = 1 - \prod_{m=1}^{12} (1 - \overline{R}_{jm})
\]  

(11)

CONCLUSIONS

The choice of the most appropriate disease frequency measure to be estimated from a NAHMS is primarily user dependent. A set of calculations have been proposed, for standardization of frequency computations. All the various estimators are derived from the incidence densities calculated from herds selected randomly and observed monthly for a period of time, in this case, 1 year. The progression of each of these measures from the incidence density (\(ID_{ijm}\)), is shown in Fig. 1. Suggested uses of each measure are also shown. Other progressions could be considered, but these were chosen based on their anticipated use and approximatable statistics.

Fig. 1. Flow chart of various disease frequency measures proposed, with definition and expected use.
Further work is needed in the areas of: (1) improved ways to calculate the $ID_{ijm}$ so as to remove bias; (2) determination of the most appropriate weighting term for stratum estimates; (3) calculation of the variances.

ACKNOWLEDGEMENTS

This project was supported in part by USDA-APHIS-VS Cooperative Agreement No. 12-16-93-229 and Michigan State University (MSU) College of Veterinary Medicine. We thank the VMOs of the Michigan Department of Agriculture, USDA-APHIS-VS, and MSU for their essential role in collecting the data. We thank the faculty and staff of the Division of Epidemiology for their input.

REFERENCES


