J. Hyg., Camb. (1972), 70, 409
Printed in Great Britain

A method for predicting proportions of affected herds from proportions of affected animals

By F. B. LEECH and R. W. M. WEDDERBURN
Rothamsted Experimental Station, Harpenden, Herts.

(*Received* 23 *December* 1971)

SUMMARY

The frequency of herds affected with 13 different diseases is shown to bear a simple relationship to the frequency of affected animals. The relationship seems to be useful for predicting proportions of affected herds.

From time to time an estimate is needed of the proportion of herds, in a population of farms, likely to contain animals affected by some disease. Given the proportion of affected animals in the population, an estimate could be obtained from the distribution of herd sizes if the relationship between population disease frequency, herd size, and proportion of affected herds were known.

We have studied published data on 14 different categories of disease in cattle and show that from a simple mathematical relationship between the three factors, the proportion of affected herds can be predicted with useful accuracy.

The data, summarized in Table 1, are taken from reports of national surveys of diseases in cattle on random samples of the farms of Britain (Leech, Davis, Macrae & Withers, 1960; Leech, Vessey & Macrae, 1964; Leech, Macrae & Menzies, 1968). The published tables show percentages of affected herds within from 4 to 8 herd-size groups. The percentages usually increase with increasing disease frequency and with increasing herd size.

For random, independent events (which diseases are not), the binomial distribution predicts that the proportion, Q, of unaffected groups of size n is q^n, where q is the proportion of unaffected individuals. Because $\log Q = n \log q$, trends relating $\log Q$ to group size are straight lines through the origin with slope equal to $\log q$. A plot of $\log Q$ against n using the data of Table 1 showed that relationships fitting the data would be curved and might miss the origin. Other transformations of Q and n were tried, but none rectified the curvature.

A plot of logit P (or logit Q) against $\log n$ showed much better promise of obtaining a reasonable fit (logit $P = \log (P/Q)$; we imply natural logarithms in both 'log' and 'logit'; note that some tables of logits use $\frac{1}{2} \log (P/Q)$). This plot suggested that either parallel or radiating straight lines might fit the points. Any one such line has the formula

$$\text{logit } P = a + b \log n$$

where a, the logit of the proportion of affected herds of one animal, might be expected to be related to logit p (p being the proportion of affected animals).

Table 1. *Data, taken from reports of national surveys, showing percentages of herds containing affected animals and population percentages of animals affected*

Disease condition	Herd size							
Calf mortality	1–5	6–10	11–20	21–30	31–40	41–60	61–80	81+
No. of herds	100	209	551	331	173	144	53	46
Percentage affected	8·0	24·9	33·1	47·7	56·6	68·8	77·4	89·1
Percentage animals affected = 3·74								

	Herd size				Percentage animals affected
	⩽ 19	20–39	40–59	60+	
No. of herds	475	482	129	72	
	Percentage affected				
Johne's disease	5·1	7·1	9·3	11·1	0·35
Summer mastitis	2·1	12·0	19·4	29·2	0·50
Grass tetany	5·1	8·1	20·2	34·7	0·57
Dystokia	21·1	29·0	28·7	44·4	1·55
Stillbirth	18·1	34·0	42·6	62·5	1·82
Acetonaemia	18·1	25·9	47·3	56·9	2·01
Abortion	24·0	39·0	47·3	63·9	2·15
Foul-in-the-foot	21·1	32·0	54·3	66·7	2·79
Acute mastitis	30·1	47·9	59·7	81·9	3·53
Milk fever	33·1	47·9	69·0	81·9	3·65
Retained placenta	32·0	57·1	71·3	79·2	4·24
Mild mastitis	44·0	58·9	66·7	81·9	6·78

	Herd size				
Udder brucellosis	⩽ 19	20–29	30–39	40–49	50+
No. of herds	599	623	390	236	414
Percentage affected	6·3	14·0	15·9	16·9	19·3
Percentage animals affected = 1·09.					

These considerations suggested that the general relationship

$$\text{logit } P_{ij} = a + b \text{ logit } p_j + (c + d \text{ logit } p_j) \log n_{ij} \qquad (1)$$

(where i represents a size-group and j a disease) should be tried, and some of the parameters fixed or eliminated to test the relative value of simpler relationships. Equation (1) implies a set of straight lines relating logit P_{ij} to $\log n_{ij}$, radiating from a point with coordinates $\log n_{ij} = -b/d$, logit $P_{ij} = a - bc/d$, and with slopes $c + d$ logit p_j.

Various parameter values were tried, using a computer program that searched for the minimum of the log-likelihood ratio by the simplex method of Nelder & Mead (1965). The log-likelihood ratio (L) was calculated from observed proportions of affected herds (P) and predicted proportions ($\hat{P}$) using the relationship

$$L = \Sigma \{R \log (P/\hat{P}) + (N - R) \log (Q/\hat{Q})\}$$

where N is the number of herds in a size-group of which R were affected; $Q = 1 - P$. This relationship implies an assumption of binomially distributed residual errors, which is contradicted by the analysis of χ^2 shown in Table 3. However, the use of

this likelihood ratio will still be justified if the errors have variances proportional, rather than equal, to binomial errors.

A series of trials with equation (1) showed:

(1) that there was no gain (in terms of the value of the log likelihood ratio per degree of freedom) from fitting more than one parameter;

(2) when b was fitted and the other parameters fixed at $a = 0$, $c = 1$, $d = 0$, the log likelihood ratio was only trivially smaller than when a was fitted and the others fixed at $b = 1$, $c = 1$, $d = 0$, which gave the equation

$$\text{logit } P_{ij} = -0{\cdot}1227 + \text{logit } p_{ij} + \log n_{ij}, \tag{2}$$

using natural logarithms, from which

$$P = \frac{0{\cdot}885\, np}{q + 0{\cdot}885\, np}. \tag{3}$$

From equation (3) P can be estimated by simple arithmetic; this seems a worthwhile advantage over the equation with b as the fitted constant, which requires logarithms for its solution.

The nature of the relationship represented by equation (3) is seen at its simplest by looking at the odds (P/Q) on a herd of size n being affected. These odds are $n(0{\cdot}885\, p/q)$. Setting s equal to the proportion within brackets, we see that the odds, ns can be represented by straight lines, with slope s, passing through the origin of a graph whose axes are P/Q, and n. However, such graphs give a misleading impression of discrepancies between observed and fitted odds that are associated with large values of P, because the statistical error in P/Q increases without limit as P approaches 100 per cent. A more realistic impression is given in Fig. 1, which shows a selection of observed points and corresponding prediction curves derived from equation (3). This figure shows, for example, that the predicted trend is less steep than the observed trend of calf deaths, but more steep than the observed trend in Johne's disease.

For other diseases, equation (3) gave predictions that also departed more or less systematically from the observed trends. We examined these discrepancies in detail (Tables 2 and 3) and conclude that they are not large enough to detract from the general usefulness of equation (3).

In Table 2, the errors are calculated as proportions of the predicted numbers of affected and unaffected herds pooled over all herd sizes. The error for udder brucellosis is shown separately; the brucellosis data were excluded from the fitting because the survey report (Leech *et al.* 1964, p. 19) comments that the examination of milk samples from individual cows was incomplete in one category of herds. The observed proportions of herds with udder brucellosis in the survey were therefore almost certainly less than the actual proportions. The other errors in Table 2 are all below 15%, which seems adequate accuracy for the purposes for which prediction might be required.

Estimates of χ^2 (Table 3), which combines the errors for affected and unaffected herds, were calculated for the total discrepancy and for discrepancies from the individual observations of Table 1. Large values of χ^2 (per degree of freedom) in

https://doi.org/10.1017/S0022172400062987 Published online by Cambridge University Press

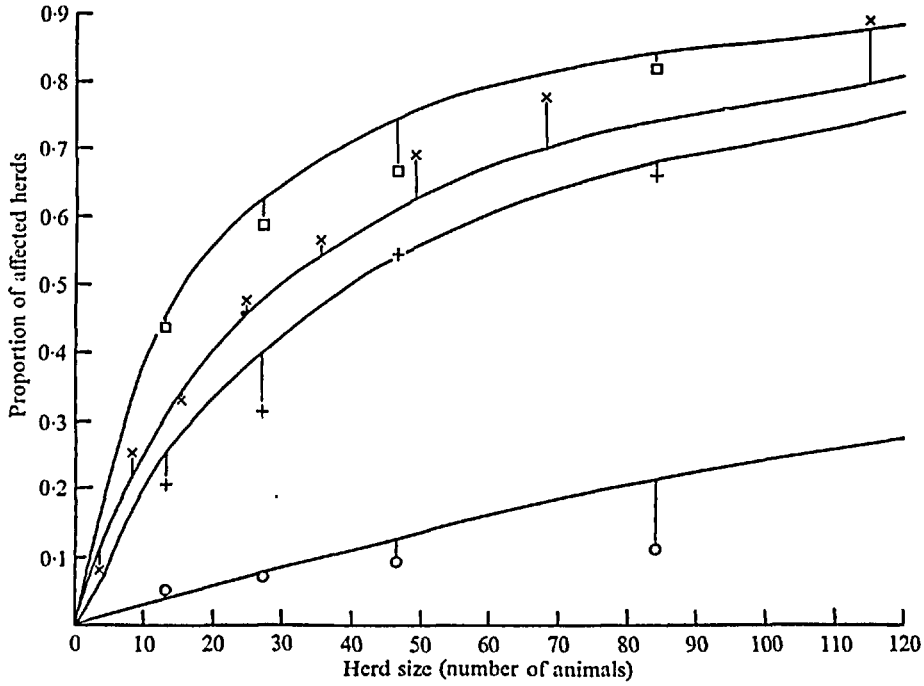

Fig. 1. Lines showing the predicted proportion of herds affected by four diseases, with the associated actual proportions. × calf deaths, ○ Johne's disease, + foul-in-the-foot, □ mild mastitis.

Table 2. *Error in estimating percentage of affected herds from* $P = 0{\cdot}885\,np/(q+0{\cdot}885\,np)$

	Discrepancy as a percentage of predicted number of herds	
	Affected	Unaffected
Calf deaths	3·4	−2·4
Johne's disease	−6·7	0·5
Summer mastitis	−4·3	0·5
Grass tetany	−14·6	1·9
Dystokia	4·7	−1·6
Stillbirth	6·6	−2·6
Acetonaemia	−10·9	4·8
Abortion	10·8	−5·1
Foul-in-the-foot	−14·0	8·3
Acute mastitis	3·0	−2·2
Milk fever	6·3	−4·9
Retained placenta	4·9	−4·4
Mild mastitis	−6·1	8·6
Udder brucella	−41·9	12·8

https://doi.org/10.1017/S0022172400062987 Published online by Cambridge University Press

Table 3. *Error* (χ^2) *in estimating percentage of affected herds from* $P = 0{\cdot}885\,np/(q+0{\cdot}885\,np)$

Disease	For the overall discrepancy (1 d.f.)	Summed over discrepancies of herd-size groups	
		d.f.	χ^2
Calf deaths	1·29	8	7·06
Johne's disease	0·42	4	8·73
Summer mastitis	0·24	4	11·84
Grass tetany	3·22	4	8·96
Dystokia	0·87	4	19·26
Stillbirth	1·99	4	4·33
Acetonaemia	6·02	4	11·08
Abortion	6·33	4	8·72
Foul-in-the-foot	13·46	4	18·57
Acute mastitis	0·77	4	3·33
Milk fever	3·53	4	7·73
Retained placenta	2·51	4	11·04
Mild mastitis	6·07	4	9·65
Total	46·72 (13 d.f.)		130·30 (56 d.f.)
Udder brucella	121·04 (1 d.f.)		144·31 (5 d.f.)

the second column of Table 3 associated with small values in the first column, show where the shape of the observed curve differed considerably from that of the predicted curve.

Although the discrepancies for calf deaths in Fig. 1 look systematic, the values of χ^2 in Table 3 show that they were of the size that would be associated with binomial error; it therefore seems more reasonable to attribute them to sampling error than to systematic departure from the model. The χ^2 for Johne's disease dicrepancies would be exceeded only in about 6 % of random samples. The survey showed that this disease was much more frequent in Channel Island than in other breeds and that the herds of Channel Island cattle were relatively small. This observation provides a sensible explanation of the systematic departure from the model. It suggests also that if data classified into Channel Island versus other breeds existed, equation (3) would give good individual predictions for the two breed groups. Breed differences could also account for the discrepancies from the predicted curve for foul-in-the-foot. In general, when the frequency of disease per animal depends greatly on factors associated with considerable differences in herd size, systematic discrepancies from a curve calculated from the average frequency per animal are to be expected. Grass tetany was about four times more frequent (per animal) in herds in the north of Scotland (averaging 41 cows) than in herds in the south-west of England (averaging 24 cows). The national average frequency of grass tetany therefore tends to overestimate the proportion of affected small herds and to underestimate the proportion of affected large herds. In such conditions, separate predictions for different geographical areas, using the regional proportions of affected animals, should be used when accurate estimates are required.

DISCUSSION

For our purposes, an empirical model seems better than a theoretical model such as the negative binomial, partly because the empirical model is simpler and partly because no single theoretical assumption about the associations between occurrences seems appropriate when such a range of diseases is being considered.

The relationship (equation (3)) has been fitted only to data for diseases in herds of cattle in Britain. Data giving both the frequency per animal and the frequency of affected herds in the same population are uncommon. We have used all the data we could find. If the same relationship is found adequate for describing the situation in other species and other countries, its general usefulness will be enhanced.

Field observations determine a proportion of affected animals much more precisely than a proportion of affected herds. This is partly because the number of herds per size group is necessarily small relative to the numbers from which the average frequency per animal is calculated. Because the frequency of affected herds in a group covers a range of herd sizes, it may be a slightly biased estimate of the frequency for the mean herd size of the group. Furthermore, mean herd size could not always be calculated precisely from the published data, and the use of some approximations may have introduced extra divergences from the relationships in the original observations.

The prediction errors in applying equation (3) to our data are presented as values of χ^2. It is clear from these that we cannot assume binomial errors for the proportion of herds infected. At least in part, this is because some factors closely associated with variation in the frequency of some diseases were unequally distributed among herd size groups.

A relationship such as equation (3) will probably be useful when the predicted proportion of affected herds is not close to zero or 100%. The useful range is affected by herd size and by disease frequency. Very large herds are expected to have at least one animal affected by any disease that is of economic importance in the population to which it belongs. Very common diseases are expected to occur in almost all herds.

REFERENCES

LEECH, F. B., DAVIS, MURIEL E., MACRAE, W. D. & WITHERS, F. W. (1960). *Disease, wastage and husbandry in British dairy herds: Report of a national survey in* 1957–58. London, H.M.S.O.

LEECH, F. B., VESSEY, M. P. & MACRAE, W. D. (1964). Brucellosis in the British dairy herd. *Animal Disease Surveys: Report No.* 4. London: H.M.S.O.

LEECH, F. B., MACRAE, W. D. & MENZIES, D. W. (1968). Calf wastage and husbandry in Britain, 1962–63. *Animal Disease Surveys: Report No.* 5. London: H.M.S.O.

NELDER, J. A. & MEAD, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308–13.