THE USE OF THE LOG-LINEAR MODEL TO EVALUATE HERD FACTORS AS DETERMINANTS OF *STAPHYLOCOCCUS AUREUS* AND *STREPTOCOCCUS AGALACTIAE* MASTITIS OCCURRENCE IN CALIFORNIA DAIRY HERDS IN 1977

Margaret A. Thorburn

In spite of the many advances made in mastitis control techniques, this costly dairy disease today remains highly prevalent worldwide. Reductions in levels of mastitis have been partially offset by changes in dairy management which tend to promote its spread (Jasper, 1980). Recently, studies have begun focusing on the identification of determinants of mastitis prevalence at the herd level. Various bulk tank milk indices have been used to identify mastitis infected or problem herds. Goodhope et al. (1980) used the bulk tank milk gel index to identify case and control herds in a study of mastitis in Ontario dairy herds. They found that case herds were less likely to raise their own replacements and cull low producers or to use teat dips, dry cow treatments or regular veterinary service. Hoare et al. (1979), in a study of relationships between mastitis prevalence at the herd level and dairy farm practices, used high bulk tank milk cell counts to identify problem herds. In a case-control study among California dairy herds, Thomas et al. (1981) analyzed bulk tank milk mycoplasmal culture results and found an association of mycoplasma with large herd size. They also found an association with high culling rates which may imply an increased introduction of potentially infected replacements or a response to the marked agalactiae often associated with mycoplasmal mastitis.

This paper examines associations among *S. agalactiae* and/or *S. aureus* infected and non-infected herds and three herd factors: herd size, location and participation in the California Dairy Herd Improvement Association (CDHIA). Log-linear model methodology (Bishop et al., 1975; Feinberg, 1980) was used to analyze the cross-classification contingency table formed by these five variables. Large multidimensional contingency tables present special problems of analysis and interpretation. Standard methods of analyzing various two-dimensional tables fail when other variables are included as additional dimensions. Marginal totals no longer represent the

---

Department of Epidemiology and Preventive Medicine, School of Veterinary Medicine, University of California, Davis, CA 95616

relationships between a pair of categorical variables, nor do they allow for the simultaneous examination of pairwise relationships or for the examination of three-factor or higher-order interactions (Feinberg, 1980). Log-linear models do not have these shortcomings.

MATERIALS AND METHODS

Data Collection: Data were obtained and linked from three sources. Herd size and location information were taken from the California Dairy Cattle Data Base, which uniquely identifies each of the 2875 dairies in California. Herds participating in CDHIA were identified from the 1977 year-end summary for CDHIA herds, obtained from the Cooperative Dairy Extension, University of California, Davis. The bacteriological culture results of a 1977 statewide survey of bulk tank milk, (Jasper et al. 1979) were used to identify S. aureus and S. agalactiae infected herds. This study covers five of the six California State Bureau of Animal Health (BAH) Disease Control Districts; northern California (district 1) is excluded because of poor coverage in the 1977 survey. Complete data were available for 1782 commercial dairies.

Category Description: The criterion for herd infection with either pathogen was isolation of that pathogen from the herd's bulk tank milk sample; non-infected herds for either pathogen were those with negative

Table 1. Methods of numerization and marginal totals for factors related to S. aureus and S. agalactiae occurrence in Calif., 1977.

Herd factor	Level of categorization	Method of numerization	Number of herds in factor levels
1. CDHIA	No participation	1	1023
	Participation	2	759
2. Herd size	10-250 cows	1	917
	250 or more cows	2	865
3. Location	District 2	1	86
	District 3	2	220
	District 4	3	607
	District 5	4	538
	District 6	5	331
4. S. aureus	Positive herd	1	1269
	Negative herd	2	513
5. S. agalactiae	Positive herd	1	1043
	Negative herd	2	739

bulk tank milk samples for that pathogen. Herd location was classified into five groups, BAH districts 2,3,4,5 and 6, and CDHIA participation into participants (official and non-official) and non-participants. Herd size was classified into two groups, 10-250 cows (small) and 251 or more cows (large). Two factors influenced the choice of herd size categories. Intervals which minimized zero elementary cells were considered important for the analysis. Consideration was also given to herd size intervals which reflected natural divisions in management methods dictated by economies of scale. Table 1 explains the treatment of the data and lists the marginal totals for each factor.

Method of Analysis: The BMDP3F computer program (Dixon and Brown, 1979) was used to form and analyze the five-dimensional contingency table. This program fits log-linear models belonging to a hierarchical set in which higher-order terms may be included only after inclusion of the related lower-order terms (see Feinberg, 1980). In order to limit the number of effects to be considered in the final model, the relative importance of each possible effect was screened using tests of marginal and partial association (Brown, 1976). These test the null hypothesis that the effect of interest does not contribute significantly to the overall fit of a model, given a certain set of effects already in that model. After indentifying necessary and questionable effects, possible log-linear models were specified and fitted.

The BMDP3F program uses an iterative proportional fitting process to generate the maximum likelihood estimates of the expected cell frequencies of the specified model(s). The goodness-of-fit of each model was tested using the likelihood-ratio chi-square statistic, defined as

$$G^2 = 2\Sigma_{ijk\ell m} x_{ijk\ell m} \ln(x_{ijk\ell m} / m_{ijk\ell m})$$

where $x_{ijk\ell m}$ equals the observed frequency in cell i,j,k,ℓ,m and $m_{ijk\ell m}$ equals the estimated frequency of cell i,j,k,ℓ,m. G^2 is distributed under the null hypothesis as a central χ^2 with degrees of freedom equal to the number of cells in the table minus the number of parameters fitted. G^2 is additive under partitioning for nested models (see Bishop et al., 1975).

A nested hierarchy of the fitted models was formed and the difference in G^2 between each pair of models was calculated. This difference is a test of the additional effects in the "higher" model conditional on the effects in the "lower" model; it is asymptotically distributed as with degrees of freedom equal to the difference in degrees of freedom for each model. The selected log-linear model was that in which additional effects did not contribute significantly, at the 0.05 level, to the fit of the model.

The effect parameter estimates (u-terms) of the selected model were calculated by the BMDP3F program. Calculations are similar to those in a factorial analysis of variance.

Proceedings of the 3rd International Symposium on Veterinary Epidemiology and Economics, 1982
Available at www.sciquest.org.nz

RESULTS

The marginal and partial association screening tests were both significant ($p<0.05$) for all five main effects, all possible first-order effects (two-way interactions) except S. aureus-herd size and S. aureus-CDHIA participation, and for the three-way interaction between the herd factors (herd size, location and CDHIA participation). These effects were therefore considered necessary for inclusion in the final model. Only the partial test statisics were significant for the S. aureus-herd size and the S. aureus-CDHIA interactions, so these two effects were considered as possibly needed in the final model. Results for all other interaction effects were non-significant.

Using these guidelines, three models were selected and fitted to the data. These models and their corresponding G^2 goodness of-fit-test statistics are listed in Table 2. Model (a) contains all main effects, all possible first-order effects and the three-way herd factor interaction. Model (b) is nested in model (a) and differs from it only by excluding the S. aureus-CDHIA participation interaction. Model (c) is nested in model (b) and differs from it only by excluding the S. aureus-herd size interaction. All three models provide good fits to the data (non-significant G^2 statistics).

The "best" of the three models was found by partioning the G^2 statistics, as shown in Table 2. It was found that neither of the first-order interactions, S. aureus-herd size and S. aureus-CDHIA,

Table 2. Tests of fit and partitioning of the likelihood-ratio statistics of models selected from screening effects of CDHIA participation (1), herd size (2), location (3), S. aureus (4) and S. agalactiae (5) infection status in California, 1977.

Model abbreviation[#]	df	G^2	Effect removed	Partitioned df	Partitioned G^2	Partitioned Prob[*]
a) 14,24,34,15,25,35,45,123	45	38.7[+]				
			14	1	1.8	0.18
b) 24,34,15,25,35,45,123	46	40.5[+]				
			24	1	2.1	0.16
c) 34,15,25,35,45,123	47	42.6[+]				

[#] The abbreviated notation describes models by means of the highest-order effects present. Inclusion of a higher-order effect requires inclusion of all related lower-order effects.

[*] Levels of significance are determined from the χ^2 distribution.

[+] The G^2 statistic for each fitted model is non-significant ($p>0.05$), implying a good fit to the data.

contributed significantly to the fit of the models lacking only these effects. Model (c), which lacks both these effects, was therefore chosen as the "best" log-linear model for describing the data. This model indicates that herd S. agalactiae infection status is significantly associated with herd size, location, CDHIA participation and S. aureus infection status, while herd S. aureus infection is significantly associated only with herd location and S. agalactiae status. Furthermore, these first-order interactions are mutually independent; that is, each two-way interaction involving S. aureus or S. agalactiae does not differ significantly in either magnitude or direction of effect over different levels of the other variables. In addition, there is a three-way interaction among the herd factors.

Let 1(i) = CDHIA participation; 2(j) = herd size; 3(k) = location 4(ℓ) = S. aureus infection status and 5(m) = S. agalactiae infection status. The fitted model (a) may be written as:

$$\ln m_{ijk\ell m} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{4(\ell)} + u_{5(m)} + u_{12(ij)} + u_{13(ik)} + u_{15(im)} + u_{23(jk)} + u_{25(jm)} + u_{34(k\ell)} + u_{35(km)} + u_{45(\ell m)} + u_{123(ijk)}$$

Table 3. Some estimated first-order effect parameters (u-terms) in the log-linear model used to describe the occurrence of S. aureus and S. agalactiae infected herds in California, 1977.

Factor interacting with S. aureus	Estimated value[a] of u-term	Factor interacting with S. agalactiae	Estimated value[a] of u-term
Location		CDHIA participation	
District 2: $u_{34(11)}$	-.110	non-participant: $u_{15(11)}$	.108
District 3: $u_{34(21)}$	-.118	participant : $u_{15(21)}$	-.108
District 4: $u_{34(31)}$	-.194		
District 5: $u_{34(41)}$	.008	Herdsize	
District 6: $u_{34(51)}$	.414	10-250 cows: $u_{25(11)}$	-.088
		251+ cows: $u_{25(21)}$	.088
S. agalactiae			
pos. herd: $u_{45(11)}$	.323	Location	
neg. herd: $u_{45(12)}$	-.323	District 2: $u_{35(11)}$	.270
		District 3: $u_{35(21)}$	-.470
		District 4: $u_{35(31)}$	.037
		District 5: $u_{35(41)}$	.089
		District 6: $u_{35(51)}$	.012

[a] Parameter values are reported for positive levels (ℓ=1; m=1) of S. aureus and S. agalactiae. Values for negative levels (ℓ=2; m=2) of each can be obtained by changing the sign of the estimated u-term.

where

$$\Sigma_i u_{1(i)} = \ldots = \Sigma_m u_{5(m)} = 0$$

$$\Sigma_i u_{12(ij)} = \Sigma_j u_{12(ij)} = \ldots = \Sigma_m u_{45(\ell m)} = 0$$

$$\Sigma_i u_{123(ijk)} = \Sigma_j u_{123(ijk)} = \Sigma_k u_{123(ijk)} = 0$$

$$i = 1,2;\ j = 1,2,3,4,5;\ k = 1,2;\ \ell = 1,2;\ m = 1,2$$

Estimates of some of the effect parameters (u-terms) for this model are listed in Table 3. Because of space limitations, all 90 estimated u-terms are not included. However, the first-order terms involving S. aureus and S. agalactiae are listed.

DISCUSSION

The primary purpose of this study was to identify herd factors associated with S. aureus and S. agalactiae occurrence on California dairy farms. Using log-linear model analysis a model was fitted which adequately described the data and clarified the relationships among the underlying variables.

The u-term estimates in Table 3 indicate the direction and relative magnitude of the first-order effects involving either of the pathogens. These associations must be examined for their biological meaning. It is generally felt that S. agalactiae is a major udder pathogen in "poorly managed" herds and only rarely a problem in other herds. In France, Plommet and Le Louedec (1975) related herd S. agalactiae problems to poor hygiene, inadequate follow-up and improperly applied therapy. Table 3 indicates that herd S. agalactiae infection is associated with large herd size and CDHIA non-participation. Either of these herd factors may indirectly reflect several factors of management which independently or in combination influence the risk of S. agalactiae mastitis. In addition, since treatments for S. agalactiae rarely cure all cows, success in eradication varies inversely with herd size (Plommet and Le Louedec, 1975). Therefore a more direct effect may be attributed to herd size.

This study failed to demonstrate an association of herd S. aureus infection with herd size, although Schalm et al. (1971) claimed that the difficulty of controlling staphylococcal mastitis increases in larger herds. Perhaps the classification of "small" herds as those with 10 to 250 cows masked this effect. However, in California, with an average herd size above 200 cows, it would be of little value to study the small herds of 12-15 cows described by Schalm et al. (1971).

S. aureus also showed no significant association with CDHIA participation. S. aureus mastitis is much more difficult to "manage" out of a herd than is S. agalactiae mastitis, primarily because of its poor response to therapy and its ubiquitous nature (Jasper, 1980). In

fact, S. aureus often replaces S. agalactiae in well managed herds (Schalm et al., 1971). If indeed, CDHIA participation is a proxy for other management factors, the lack of association with herd S. aureus infection is not surprising.

S. aureus herd infection is however associated with herd location. According to the u-term estimates in Table 3, dairies in district 6 (southern California) are strongly associated with S. aureus positive bulk tank milk. Previous studies have shown strong associations between S. aureus udder infection and various housing and other environmental factors (Bakken, 1981). The southern California milk shed is highly industrialized, with very large herd size dairies concentrated in a small land area. There are many environmental features associated with the district 6 dairies which may be confounded with the S. aureus-location interaction. Similarly, the S. agalactiae-location interaction may reflect confounding environmental, and possibly managerial, factors.

As shown in Table 3, a positive association exists between positive bulk tank milk for the two pathogens (and, conversely, a positive association exists between negative bulk tanks for both). As mentioned earlier, S. aureus often replaces S. agalactiae following eradication of the latter from a herd. The current findings, however, appear to contradict this observation. But, because of the cross-sectional nature of this study, it is difficult to comment on this association.

Herd size, location and CDHIA participation are not only associated with the occurrence of mastitis, but also with the prevalence of other possible mastitis determinants. Some of these determinants have been mentioned above; others would include specific aspects of a dairies milking machine management, level of automation, herd health program, etc. A correct evaluation of the role of any of these latter factors in mastitis occurrence first requires adjustment for the confounding effects of the herd factors included in this study. Specifically, in California studies at the herd level of S. aureus mastitis must adjust for herd location and studies of S. agalactiae mastitis must adjust for herd size, location and CDHIA participation. In addition, studies of either pathogen should consider the occurrence of the other.

Log-linear model methodology allows for the identification of significant associations among disease problems and possible disease determinants. Once such associations are identified, they can be specified in the final fitted model. The log-linear model automatically adjusts for the distribution of cases across the specified variables and interactions of variables. This technique should prove useful in any study attempting to identify disease determinants of a discrete nature.

REFERENCES

1. Bakken, G. 1981. Environment and bovine udder diseases in the loose housing system for dairy cows with reference to relevant data from the cowhouse system. Acta Agric. Scand. (in press).

2. Bishop, Y.M., S.E. Feinberg and P.W. Holland. 1975. Discrete Multivariate Analysis: Theory and Practice. MIT Press, Cambridge, Mass. 557 p.

3. Brown, M.B. 1976. Screening effects in multidimensional contingency tables. Appl. Statist. 25: 37-46.

4. Dixon, W. J. and M.B. Brown (Editors). 1979. BMDP-79 Biomedical Computer Programs. University of California Press, Berkeley. 880 p.

5. Feinberg, S.E. 1980. The Analysis of Cross-classified Categorical Data. 2nd ed. MIT press, Cambridge, Mass. 198 p.

6. Goodhope, R.G. and A.H. Meek. 1980. Factors associated with mastitis in Ontario dairy herds: a case-control study. Can. J. Comp. Med. 44: 351-357.

7. Hoare, R.J.T., D.J. Critchley, E.B. Dettman, R.F. Sheldrake and L.R. Fell. 1979. Mastitis control: a survey of farm practices and their relation to bulk milk cell counts. Aust. J. Dairy Tech. 34: 91-96.

8. Jasper, D.E. 1980. Mastitis. Ch. 23 in: Bovine Medicine and Surgery. H.E. Amstutz (Editor). American Veterinary Publications, Santa Barbara, Calif., pp. 1049-1092.

9. Jasper, D.E., J.D. Dellinger, M.H. Rollins and Hakanson, H.D. 1979. Prevalence of mycoplasmal bovine mastitis in California. Am. J. Vet. Res. 40: 1043-1047.

10. Plommet, M. and Le Louedec, C., 1975. The role of antobiotic therapy during lactation in the control of clinical and sub-clinical mastitis. In: F.H. Dodds, T.K. Griffen and R.G. King (Editors), Proceedings of the International Dairy Federation Seminar on Mastitis Control, April, 1975, at Reading University, England. IDF, Bruxelles, p. 669-678.

11. Schalm, O.W., E.J. Carroll and N.C. Jain. 1971. Bovine Mastitis. Lea and Febiger, Philadelphia, Penn. 360 p.

12. Thomas, C.B., P. Willeberg and D.E. Jasper. 1981. Case-control study of bovine mycoplasmal mastitis in California. Am. J. Vet. Res. 42: 511-515.