IMPACT OF THE ANALYTICAL TREATMENT OF EXPLANATORY VARIABLES ON THE EXPLANATORY/RESPONSE RELATIONSHIP

THOMAS B. FARVER AND RUBEN N. GONZALEZ*

The assessment of the relationship between an outcome or response variable and one or more risk factors is a common analytical task of researchers in veterinary epidemiology. This assessment frequently involves adjustment for other known risk factors or nuisance information such as age, sex, breed, weight, lactation, level of milk production, etc. using a multiple regression-type analysis. Frequently researchers choose to categorize on variables such as age, lactation, and milk production often deciding upon 2 or 3 categories. The present work was undertaken to demonstrate the consequences of not making full use of the information available in a data set.

METHODS

This study made use of data that was generated from a cohort study designed to determine the efficacy of a mutant Escherichia coli vaccine in preventing gram-negative mastitis in dairy cows (Gonzalez, et al., 1987). The present study made use of information obtained on 3 variables: mastitis status (whether or not a cow developed clinical coliform mastitis), vaccination status (whether or not a cow was vaccinated with an E. coli J5 vaccine), and lactation (second, third,..., ninth). Table 1 gives the crosstabulation of the 460 cows in the study by mastitis status (M), vaccination status (V) and lactation (L). Mastitis status was considered the response variable and vaccination status and lactation as explanatory variables. The goal of an analysis involving these variables would be to determine the nature of the association, if any, between the explanatory variables and the response variable. The objective of this present study was to demonstrate how the analytical treatment of lactation, an ordinal variable, can have an impact on what is concluded regarding this association. To this end, the data were analyzed 4 times treating lactation each time in a different way. In the first analysis, cows were categorized into one of 3 lactation groups, lactations 2 and 3, lactations 4, 5, and 6, and lactations 7, 8, 9, and lactation was treated as a nominal variable. The second analysis also used the 3 lactation categories but treated lactation as an ordinal variable making use of the ordering to these categories. In the third analysis, each lactation was treated as a separate category and lactation was treated as a nominal variable. The fourth analysis used 8 lactation categories but treated lactation as an ordinal variable. The data were analyzed using loglinear model methodology (Fienberg, 1980). Information on the ordered structure of the lactation categories was incorporated into the second and fourth analyses using the method given in Fienberg (1980). The computer software used to perform the analyses were BMDP (Dixon, 1985) and SPSS^{X} (Nie, et al., 1975).

---

*School of Veterinary Medicine, University of California, Davis (USA), 95616

Proceedings of the 5th International Symposium on Veterinary Epidemiology and Economics, 1988
Published in Acta Veterinaria Scandinavia, Supplementum 84
Available at www.sciquest.org.nz

Table 1. Crosstabulation of 460 dairy cows by vaccination (with an E. coli J5 vaccine) status, clinical coliform mastitis status and lactation.

Vaccination Status	Mastitis Status	Lactation 2	3	4	5	6	7	8	9
Unvaccinated	No Mastitis	82	61	19	16	11	8	2	2
Unvaccinated	Mastitis	5	10	2	5	4	1	1	1
Vaccinated	No Mastitis	78	65	29	25	18	7	2	0
Unvaccinated	Mastitis	3	1	0	1	0	0	0	1

RESULTS

Three lactation groups treated nominally. Screening methods identified the model L,VM as the "best-fitting" loglinear model with a goodness-of-fit likelihood-ratio statistic that was nonsignificant (p=.154). Using the logit model corresponding to the "best-fitting" loglinear model, the estimated odds of a cow having mastitis (relative to not having mastitis) when not vaccinated was .144 while the odds of a cow having mastitis when vaccinated was .027. The ratio of these odds, the odds ratio $\theta_{21k:22k}$, was 5.4 (95% CI: 2.2,13.2) meaning that cows that were not vaccinated had a "risk" for developing mastitis which was 5.4 times greater than that for cows that were vaccinated. It should be noted that because there was found no statistically significant association between lactation and mastitis status, the logit model is not dependent on k and the odds ratio holds for any lactation groups assumed.

Three (3) lactation groups treated as ordinal categories. The mastitis status x linear lactation term was nearly significant (p=0.076) and was included in the logit model along with the interaction between mastitis status and vaccination status. Thus for a constant lactation, the ratio of odds of a cow having mastitis when not vaccinated to the odds of a cow having mastitis when vaccinated, $\theta_{21k:22k}$, was 5.4 (95% CI: 2.2,13.2). For constant vaccination status, the ratio of odds of a cow in lactation group 2 (lactations 4, 5, and 6) having mastitis to the odds of a cow in lactation group 1 (lactations 2 and 3) having mastitis, $\theta_{2j2:2j1}$, was 1.6 (95% CI: .97,2.7). Also for constant vaccination status, the ratio of odds of a cow in lactation group 3 (lactations 7, 8, and 9) having mastitis to the odds of a cow in lactation group 1 having mastitis, $\theta_{2j3:2j1}$, was 2.7 (95% CI: .95,7.6).

Each lactation treated as a separate category. Screening methods identified the model VM,LM as the "best-fitting" loglinear model demonstrating excellent fit (p=.433). Using the logit model corresponding to the loglinear model VM,LM, for a constant lactation, the ratio of the odds of a cow having mastitis when not vaccinated to the odds of a cow having mastitis when vaccinated was 5.5 (95% CI: 2.2,13.7). With vaccination status held constant, the ratio of the odds of a cow in lactation 3 having mastitis to the odds of a cow in lactation 2 having mastitis, $\theta_{2j3:2j2}$, was 1.8 (95% CI: .68,4.6). Odds ratios obtained for cows in other lactations (with lactation 2 as the reference category and for constant vaccination status) were: $\theta_{2j4:2j2}$=.96 (95% CI: .19, 4.8), $\theta_{2j5:2j2}$=3.4 (1.1,10.7), $\theta_{2j6:2j2}$=3.2 (.86,7.5), $\theta_{2j7:2j2}$=1.3 (.14,11.1), $\theta_{2j8:2j2}$=4.8 (.44,51.9), and $\theta_{2j9:2j2}$=17.5 (1.9,157).

Proceedings of the 5th International Symposium on Veterinary Epidemiology and Economics, 1988
Published in Acta Veterinaria Scandinavia, Supplementum 84
Available at www.sciquest.org.nz

Eight (8) lactation groups treated as ordinal categories. The mastitis status x linear lactation term was highly significant (p<.0001) and was included in the logit model along with the interaction between mastitis status and vaccination status. Thus for a constant lactation, the ratio of odds of a cow having mastitis when not vaccinated to the odds of a cow having mastitis when vaccinated, $\theta_{21k:22k}$, was 5.4 (95% CI: 2.2,13.2). With vaccination status held constant, the ratio of the odds of a cow in lactation 3 (k=3) having mastitis to the odds of a cow in lactation 2 (k=2) having mastitis, $\theta_{2j3:2j2}$, was 1.3 (95% CI: 1.1,1.6). Odds ratios obtained for cows in other lactations (with lactation 2 as the reference category and for constant vaccination status) were: $\theta_{2j4:2j2}$=1.7 (95% CI: 1.1,2.4), $\theta_{2j5:2j2}$=2.2 (1.3,3.8), $\theta_{2j6:2j2}$=2.8 (1.3,5.9), $\theta_{2j7:2j2}$=3.6 (1.4,9.1), $\theta_{2j8:2j2}$=4.6 (1.5,14.2), and $\theta_{2j9:2j2}$=6.0 (1.6,22.1).

DISCUSSION

These results demonstrate very nicely that the analytical treatment of a variable can have a pronounced impact on the results of an analysis. Two issues that have been identified as being important are method of categorization and whether or not the ordinal structure of the categories has been taken into consideration. When three lactation categories formed by aggregating cows in lactations 2 and 3, lactations 4, 5, and 6, and lactations 7, 8, and 9 were used without taking into account the ordinal nature of lactation, no effect of lactation on mastitis was found. However, when each of the 8 lactations were used as separate categories, again without taking into account the ordinal nature of lactation, animals in lactations 5, 6, 8 and 9 were shown to be at a higher "risk" of having mastitis compared to cows in lactation 2 (vaccination status held constant). Furthermore even when 3 lactation categories were used but the ordinal structure of these lactation categories was taken into account, a lactation effect was observed. Both of the higher lactation groups were shown to be at a higher "risk" of having mastitis compared to the lowest lactation group (vaccination status held constant) with both estimated odds ratios being nearly statistically significant. This latter result became more pronounced when 8 lactation categories were used and the ordinal nature of lactation taken into account. These findings seem to indicate that in those cases where it is necessary to categorize on an explanatory variable one needs to be very careful in evaluating the influence of the nature of the categorization and where appropriate make use of the ordinal nature of the categories.

REFERENCES

Dixon, W.J., ed. (1985) BMDP Statistical Software 1985. University of California Press, Berkeley, Calif. 725 p.

Fienberg, S.E. (1980) The Analysis of Cross-Classified Categorical Data. 2d ed. The MIT Press, Cambridge, Mass. 198 p.

Gonzalez, R.N., Cullor, J.S., Jasper, D.E., Farver, T.B., Bushnell, R.B. Efficacy and safety of a meat-killed J5 Escherichia coli vaccine to prevent post-calving clinical mastitis in dairy cows. Presented at the Conference of Research Workers in Animal Disease, Chicago, November, 1987.

Nie, N.H., Hull, C.H., Jenkins, J.G., Steinbrenner, K. and Bent, D.H. (1975) Statistical Package for the Social Sciences. McGraw-Hill, New York, N.Y.

Proceedings of the 5th International Symposium on Veterinary Epidemiology and Economics, 1988
Published in Acta Veterinaria Scandinavia, Supplementum 84
Available at www.sciquest.org.nz