Epidémiol. santé anim., 1997, 31-32

CAUSAL MODELLING WITH AND WITHOUT RANDOM EFFECTS - AN EXAMPLE FROM A DEER HERD HEALTH AND PRODUCTIVITY PROFILING STUDY

Audigé L.[1], Pfeiffer D.U.[2], Morris R.S.[2], Wilson P.R.[2]

De nombreuses bases de données analysées par des épidémiologistes vétérinaires sont sujet au problème de l'effet troupeau. Dans le cadre du développement de modèle de causalité, l'utilisation de modèles statistiques tenant compte de cet effet a été préconisée sous peine de conduire à des inférences statistiques erronées.
Dans le cadre d'une enquête longitudinale dans 15 élevages de cerfs en Nouvelle Zélande, de nombreuses données ont été collectées au niveau de l'élevage, du groupe d'animaux (comme par exemple le harem) et de l'individu. En utilisant la technique des pistes associées à des analyses de régression logistique ordinaire, une série d'hypothèses a été proposée quant aux relations entre facteurs de causalité possibles d'une part et les performances de reproduction d'autre part. Dans cet article, quelques bases de données concernant la reproduction des cerfs ont été réanalysées en tenant compte d'un possible effet troupeau. Par exemple, un modèle logistique a été développé pour expliquer la conception des biches adultes pendant le rut. Le modèle final a montré l'importance significative de 3 caractéristiques individuelles des biches, et de 2 facteurs liés à la conduite des harems. Après prise en compte d'un effet troupeau au niveau des harems, seuls les deux facteurs relatifs aux harems n'étaient plus significatifs. Le coefficient de régression d'un facteur individuel a changé.
Ceci démontre que les résultats de modélisation doivent être interprétés avec prudence en présence d'effet troupeau. L'utilisation de modèles statistiques classiques reste valable dans le cadre d'études exploratrices visant au développement d'hypothèses. L'utilisation de modèles permettant de déceler l'existence d'un effet troupeau est toutefois utile pour une meilleure exploration des données, et nécessaire pour une finalisation des modèles de causalité.

INTRODUCTION

Many data sets which are being analysed by veterinary epidemiologists are subject to clustering effects. One typical example is the situation where data on individual animals is being analysed with multiple animals belonging to the same herd. In this case there could be clustering due to a herd effect. When developing causal models based on such data sets, ignoring the clustering can result in underestimation of the standard errors of regression parameter estimates as well as incorrect regression coefficient estimates. Over recent years statistical techniques have been developed for regression models which allow taking this effect into account through inclusion of random effect terms (McDermott et al., 1994).
In this paper several data sets from a deer herd health and productivity profiling study which was conducted in New Zealand (Audigé et al., 1994; Audigé, 1995) have been used to discuss the effects and rationale of using different statistical modelling approaches.

MATERIALS AND METHODS

A two-year longitudinal observational study was carried out on 15 red deer farms beginning in March 1992 (Audigé et al., 1994; Audigé, 1995). This study has provided numerous path diagrams using standard multivariable logistic and linear regression analyses, which allow generation of hypotheses on the potential effect of various risk factors on the most important health and production outcomes.
Risk factors were recorded at the individual, mob or herd, and farm level. However, in a first approach, all risk factors were analysed as individual deer variables, i.e. the potential effect of clustering of animals within mobs was not taken account of. The method for data analysis has been fully described (Audigé, 1995).
During the following analyses the effect of random effect terms at farm, farm within year and mob within year level on the regression coefficient estimates were assessed for 3 different models concerning deer reproduction. Reanalysed outcome variables selected for discussion in this paper were whether adult and yearling hinds, respectively, had conceived and whether adult hinds were lactating at the time of weaning, indicating successful rearing of the offspring, i.e. dichotomous variables. The statistical software used for these analyses was SAS for Windows. The SAS macro GLIMMIX was used for the random effects logistic regression analyses.

RESULTS

A statistical model was developed to explain conception in adult hinds using individual animal level variables as well as mob management variables. The final logistic regression model included the main effects of the individual animal level variables: *Pure NZ Breed*, *Hind lactating at Weaning*, *Pre-mating Body Condition Score* and the mating mob level variables: *Use of backup stag* and *Use of at least one experienced sire stag*. Table I presents the regression parameter estimates after inclusion of random effects representing the different levels of aggregation. On introduction of the mob level random effect into the statistical analysis both mating mob level

[1] Institute of Virology and Immunoprophylaxis, P.O. Box, CH-3147 Mittelhäusern, Switzerland
[2] Department of Veterinary Clinical Sciences, Massey University, Palmerston North, New Zealand

variables become insignificant. The magnitude of the estimated coefficient for the variable *Pure NZ Breed Hind* changed considerably, whereas the other two individual animal level variables did not change much. In this case inclusion of random effect terms did markedly change the coefficients of the variables representing information at the level of the random effects term. The inferences regarding the individual animal level variable did not change to a significant extent. Comparison of the deviance estimates suggests that the model with mating mob level random effect significantly improves overall model fit.

Table I

Deviances and regression coefficient estimates expressed as odds ratios and their 95% confidence limits for the logistic regression models for conception in adult hinds based on inclusion of different random effect terms

Risk factors[1]	No random effect term	Inclusion of farm level random effect	Inclusion of farm / year level random effect	Inclusion of mating mob / year level random effect
Pure NZ Breed	2.99 (1.79-5.03)	2.37 (1.36-4.14)	2.37 (1.38-4.06)	1.81 (1.05-3.13)
Hind lactating at Weaning	3.38 (1.85-6.16)	3.41 (1.91-6.08)	3.50 (1.99-6.16)	3.18 (1.84-5.49)
Pre-mating Body Condition Score	1.86 (1.34-2.58)	1.79 (1.31-2.45)	1.75 (1.29-2.38)	1.68 (1.25-2.25)
Use of at least one experienced sire stag	2.27 (1.40-3.68)	2.16 (1.32-3.51)	2.05 (1.21-3.47)	1.75 (0.84-3.64)
Use of backup stag	1.99 (1.21-3.25)	1.75 (0.94-3.25)	1.82 (0.99-3.36)	1.80 (0.92-3.53)
Deviance	703.32	685.67	678.4	652.68

[1] All risk factors are dichotomous except *Pre-mating body condition score* (scores from 1 to 5).

The final logistic regression model for conception of yearling hinds included the individual animal level variables: *Percentage of New Zealand blood lines*, *Standardised pre-mating body weight* and *Growth rate between April 1 and June 1*. Table II presents the regression coefficients and deviance estimates for the different models. Inclusion of the different random effects terms did not result in notable changes of the coefficients or their confidence limits, except in the case of growth rate. Deviance did change considerably after inclusion of the farm level random effects term as did the coefficient for growth rate. This suggests that, to some extent, this variable represented differences between herds.

Table II

Deviances and regression coefficient estimates expressed as odds ratios and their 95% confidence limits for the logistic regression models for conception in yearling hinds based on inclusion of different random effect terms

Risk factors[1]	No random effect term	Inclusion of farm level random effect	Inclusion of farm / year level random effect	Inclusion of mating mob / year level random effect
Percentage of New Zealand blood lines	1.02 (1.01-1.03)	1.01 (0.998-1.02)	1.014 (1.002-1.03)	1.018 (1.005-1.03)
Standardised pre-mating body weight	1.05 (1.01-1.08)	1.06 (1.02-1.097)	1.058 (1.02-1.098)	1.05 (1.014-1.09)
Growth rate betw. April 1 and June 1	1.089 (1.00-1.014)	1.01 (1.004-1.016)	1.009 (1.003-1.015)	1.008 (1.0025-1.014)
Deviance	552.82	509.56	495.08	491.99

[1] Risk factors measured on a continuous scale (units are %, kg and g/d, respectively).

The logistic regression model for lactational status of adult hinds included the following individual animal risk factors: *Adult hind over 3 years old at mating, Body condition score difference between March (pre-mating) and September (post-winter), Hind conceived before May 1 (early conception)* and the mob level variables: *Mean residual (ie post-grazing) pasture sward height below 5 cm, At least one stag over 15 months old grazed with hinds during hind lactation, Mean paddock size (deer-fenced area), Average of daily maximum temperatures* and *Average daily sunshine score.* Table III shows the results for the different models. Model fit improves markedly by including the mating level random effect. The odds ratio for the age category of the adult hind at mating does seem to represent mob level variation to a limited extent. The standard errors for the variable representing residual pasture sward height and for average daily sunshine score increase significantly resulting in these variables becoming insignificant.

Table III
Deviances and regression coefficient estimates expressed as odds ratios and their 95% confidence limits for the logistic regression models for lactational status in adult hinds based on inclusion of different random effect terms

Risk factors	No random effect term	Inclusion of farm level random effect	Inclusion of farm / year level random effect	Inclusion of mating mob / year level random effect
Adult hind over 3 years old at mating	3.08 (2.21-4.28)	3.04 (2.18-4.23)	2.67 (1.9-3.76)	2.81 (2.0-3.9)
Body condition score difference[1]	2.44 (1.49-4.02)	2.41 (1.46-3.96)	2.34 (1.44-3.82)	2.41 (1.49-3.92)
Early conception	1.52 (1.21-1.91)	1.52 (1.21-1.92)	1.49 (1.18-1.89)	1.54 (1.23-1.93)
Mean pasture sward height <5 cm	0.58 (0.34-0.997)	0.59 (0.33-1.08)	0.67 (0.29-1.55)	0.61 (0.30-1.25)
Stags with hinds during hind lactation	0.19 (0.116-0.313)	0.17 (0.1-0.3)	0.11 (0.05-0.23)	0.16 (0.09-0.29)
Mean paddock size[2]	0.94 (0.91-0.96)	0.94 (0.91-0.97)	0.95 (0.91-0.99)	0.94 (0.91-0.98)
Average of daily maximum temp.[2]	0.76 (0.66-0.88)	0.75 (0.64-0.88)	0.72 (0.58-0.90)	0.74 (0.62-0.88)
Average daily sunshine score[3]	5.87 (1.48-23.3)	5.77 (1.29-25.8)	8.43 (0.92-77.6)	5.26 (0.85-32.55)
Deviance	1152.94	1148.84	1120.39	1099.81

[1] Risk factor measured on a ordinal scale (body condition score from 1 (lean) to 5 (fat)).
[2] Risk factors measured on a continuous scale (units are ha and ºC , respectively).
[3] Score on a continuous scale from 0 (never sunny) to 1 (always sunny).

DISCUSSION

In veterinary epidemiology, when analysing individual animal outcomes (e.g. occurrence of disease or performance) that have been recorded from clusters such as herds or farms, individual observations within clusters are not independent. This dependence is known to be caused by "herd effects". Herd effects result from differences in factors such as genetics, climate and management practices, known statistically as "random effects" because they are often unmeasured or unmeasurable. Herd effects may also result from intraherd correlation, i.e. when the performance of one individual animal depends on that of others in the herd. A good introduction on herd effects is discussed in Curtis et al. (1993).

The problem of herd effect can further developed using the examples of adult and yearling hind conception (Audigé, 1995), i.e. dichotomous outcomes. When using the standard logistic regression model (LR) to model individual probabilities of hinds conceiving, it was assumed that each hind exposed to the same set of risk factors has the same probability of conceiving. This assumption is not necessarily valid, as it is in fact likely, that hinds in heat stimulate other hinds within harems, thus being more likely to conceive. In such investigations, it seems necessary to use a statistical method adjusting for herd effect (McDermott et al., 1994). For categorical outcomes, as in these examples, random-effect logistic regression may be used to test for the significance of path coefficients and the presence of extrabinomial variation (Curtis et al., 1988).

In the context of this deer study (Audigé et al., 1994), the statistical adjustment of herd effect was not deemed a critical issue since the study aimed at the generation of hypotheses regarding the relative importance of risk factors for various outcomes. The current analyses however were carried out to assess the influence of clustering effects on the inferences drawn from the initial models. The results have shown specifically in the case of aggregate level variables such as herd or mob level parameters that it was very informative also to test for the presence of overdispersion. With the current data, some of the mob level parameters did appear to represent to some extent between mob and farm differences, resulting in significant changes in the coefficients and/or their confidence limits after introduction of the random effects term. The model for conception of yearling hinds only included individual animal effects, but the interrelationship between farm management and growth rate became evident through a change in the coefficient for the growth rate parameter.

REFERENCES

Audigé L., Wilson P.R., Morris R.S. and Pfeiffer D.U. (1994) Herd health and production profiling as an epidemiological tool for the study of farmed deer in New Zealand, in: The Kenya Veterinarian, Proceedings of the 7th International Symposium of Veterinary Epidemiology and Economics 18(2): 344-346

Audigé L. (1995) Deer herd health and production profiling. PhD thesis, Massey University, New Zealand, 1995, 550 pp.

Curtis C.R., Mauritsen R.H., Salman M.D., and Erb H.N. (1988) The enigma of herd: a comparison of different models to account for group effects in multiple logistic regression analysis. in: Acta Veterinaria Scandinavica, Proceedings of the 5th International Symposium of Veterinary Epidemiology and Economics, No 84 :462-465

Curtis C.R., Mauritsen R.H., Kass P.H., Salman M.D., and Erb H.N. (1993) Ordinary versus random-effects logistic regression for analyzing herd-level calf morbidity and mortality data. Preventive Veterinary Medicine, 16 :207-222

McDermott J.J.; Schukken Y.H., and Shoukri M.M. (1994) Study design and analytic methods for data collected from clusters of animals. Preventive Veterinary Medicine, 18 :175-191