

EPIDEMIOLOGICAL STUDIES OF SUB-ACUTE RECURRENT DISEASES WITH TIME DEPENDENT COVARIATES - A METHOD OF ANALYSIS FOR LONGITUDINAL STUDIES USING EQUINE LOWER AIRWAY DISEASE AS AN EXAMPLE

Wood J.L.N.¹, Newton J.R.¹, Chanter N.¹, Townsend H.G.G.², Lakhani, K.H.¹, Mumford J.A.¹

Les maladies respiratoires chez les jeunes chevaux de course sont d'importance capitale. Nous avons réalisé une étude longitudinale sur une période de 3 ans pour étudier les facteurs de risque associés à ces maladies. La définition des cas se base sur les observations cliniques et cytologiques. Les 188 chevaux inclus sont observés plusieurs fois. Au total nous disposons de 1628 unités d'observation (Cheval mois). Les données sont regroupées dans le temps, répétées sur les mêmes animaux, et les animaux sont regroupés dans plusieurs centres d'entraînement. Les données sont analysées avec trois modèles différents : régression logistique ordinaire (RLO), régression logistique avec des effets mixtes (RLEM) et la régression logistique conditionnelle (RLC) où la strate est le cheval.

*Les modèles RLO et RLEM montrent une association significative entre les maladies respiratoires et les variables suivantes : âge, saison, centre d'entraînement, infection par *S. zooepidemicus*, *Pasteurella/Actinobacillus* spp, *Mycoplasma equirhinis* et *Serratia* spp, l'apparition de la maladie et l'infection par *M. equirhinis* le mois précédent. Les trois dernières variables ne sont pas significatives avec le modèle CLR. Le modèle CLR donne des résultats divergents à cause de l'exclusion de l'analyse des strates sans événement (absence de maladie). Le modèle RLEM semble être le plus adapté pour l'analyse des données longitudinales.*

INTRODUCTION

Respiratory disease is a significant cause of days lost from training in young Thoroughbred TB racehorses (Rossdale *et al.*, 1985). Although traditionally, the aetiology of this disease has largely been assumed to be viral, investigations have shown that many outbreaks are not attributable to infection with known equine viruses, but frequently involve mixed bacterial (Wood *et al.*, 1993) and mycoplasma infections (Wood *et al.*, 1997), particularly in young animals (Burrell *et al.*, 1996).

Although most investigations of equine respiratory disease in racehorses have concentrated on outbreaks, much of the disease is in fact sub-acute and recurrent in nature (Burrell *et al.*, 1996); moreover, it clusters both in time and space (within training yard) and the causes, or associated infections, change in time. This poses analytical problems because most standard epidemiological designs are suitable for diseases that only occur once, do not cluster and have causes that do not change in time.

Some authors have used ordinary logistic regression (OLR) analyses, ignoring the lack of independence of data from each subject. However, methods of analysis suitable for longitudinal studies with repeated observations are available. Generalised estimating equations (GEE) (Liang & Zeger, 1986), mixed effects logistic regression models (MELR) with random effects (Hedeker & Gibbons, 1994) and conditional logistic regression (CLR) (Montgomery *et al.*, 1990; Burrell *et al.*, 1996) are examples.

In an analysis of data from a longitudinal study of equine respiratory disease, we compare results from analyses using OLR, CLR and MELR with random effects. It was of interest to determine effects of infections at the time of disease, as well as the effects of infections in the previous time period. Some variables were fixed at the level of horse (e.g. training yard), whereas others varied from observation to observation (e.g. infections).

MATERIALS AND METHODS

Data were collected from TB racehorses in training kept in seven different yards over a 38 month period from November 1993. Six yards were studied at all times. Ten to fifteen horses per yard were selected at random at the beginning of the study, monitored daily by staff for signs of disease and examined & sampled on a monthly basis. Horses in the study leaving the training yard were replaced at the beginning of the following year by other, randomly selected horses. Data were collected from 188 horses and after development of time lagged variables (see below) and exclusion of observations with missing values (in particular the first observation from each animal where the previous month's information was missing), the dataset comprised 1231 horse months from 158 horses. Disease occurred in 234 (14%) of months in the dataset.

The disease of interest was lower airway disease, graded on an ordinal scale of 0 to 3 on a clinical and cytological basis following endoscopic examination of the trachea to the level of the carina with transendoscopic collection of a tracheal wash sample, as previously described (Burrell *et al.*, 1996). All tracheal wash samples were collected after exercise. Lower airway disease was defined as a score ≥ 2 .

¹ Centre for Preventive Medicine, Animal Health Trust, P.O.Box 5, Newmarket, Suffolk, CB8 7DW UK

² Department of Veterinary Internal Medicine, Western College of Veterinary Medicine, University of Saskatchewan, Saskatoon, Canada

Explanatory variables of interest included age, season and viral, mycoplasma and other bacterial infections. Infections with Equine Herpesviruses 1 and 4 (EHV1/4), Equine Rhinovirus-1 (ERV-1) and Adenovirus were assessed serologically. Lower airway infections with mycoplasma and bacteria were assessed quantitatively by culture from tracheal wash samples (Wood *et al.*, 1997) and upper respiratory tract bacterial infections were also assessed by culture of nasopharyngeal swab specimens.

Simple univariate data analysis was carried out in Epi-info (Dean *et al.*, 1994). Further multivariate analyses were carried out in EGRET (SERC, 1989) after deriving explanatory variables lagged by one time period.

The unit of observation was the horse month. Repeated observations from each horse were not independent, neither were observations from within each training yard nor from within each time period. Results from standard OLR analysis, ignoring the clustering of observations, were compared to those from MELR with random effects, using the horse as the random effect, and also with results from CLR, with one stratum per horse. Variables were tested in multivariate models if they were associated with disease ($p < 0.3$) at the univariate level and were included in final models if they resulted in a significant change in Likelihood Ratio Statistic (LRS) or if they were significantly associated with the outcome, using Wald χ^2 test at a significance level of $p = 0.05$ (Hosmer and Lemeshow, 1989). Interaction terms between all main effect variables were tested. Time lagged explanatory variables were treated in the same way as other explanatory variables. Training yard was treated as a fixed effect in all analyses. We also included an autoregressive variable, PREVIS2 = presence of disease in the previous time period. Such a variable might absorb unknown effects related to the usual duration of disease and horses' susceptibility and immunity to disease.

RESULTS

Age, season, infection with *Streptococcus zooepidemicus*, *Pasteurella/Actinobacillus* spp and *Mycoplasma equirhinis* were significant in all models (Table I). In OLR and MELR, trainer, *Serratia* spp. infection, along with disease and *M. equirhinis* infection in the previous time period were also significant, although with reduced or marginal significance in MELR, but none of these were significant in CLR.

Results from OLR and MELR were similar. However, coefficient standard errors were generally larger in MELR and hence significance were smaller. In stratified, CLR analysis, outcome was concordant in 92 horses (87 of which remained healthy) and so CLR analyses were only based on data from 90 horses. Also, because horses were nested within training yard, it was not possible to estimate effect of trainer using these stratified methods.

Table I
Comparison of Results from Analysis using OLR, MELR and CLR, with all models optimised

Variable	β	OLR S.E. β	p value	β	MELR S.E. β	p value	β	CLR S.E. β	p value
Intercept	-3.198	0.496	<0.001	-3.198	0.496	<0.001	-		
Age: ≤ 2 years	referent			referent			referent		
= 3 years	-0.905	0.252	0.001	-0.939	0.288	0.001	-1.17	0.481	0.015
≥ 4 years	-0.302	0.329	0.39	-0.464	0.395	0.24	-1.01	0.783	0.198
Season: Nov-Jan	referent			referent			referent		
Feb-April	0.921	0.303	0.83	0.957	0.318	0.83	0.826	0.343	0.018
May-July	0.207	0.315	0.002	0.255	0.331	0.002	0.397	0.358	0.27
Aug-Oct	-0.146	0.352	0.51	-0.175	0.370	0.51	-0.419	0.413	0.31
Trainer: 1	referent			referent			-		
2	-0.169	0.501	0.74	-0.078	0.577	0.89			
3	-0.930	0.609	0.13	-0.880	0.679	0.20			
4	0.193	0.505	0.70	0.297	0.584	0.61			
5	-0.392	0.533	0.46	-0.372	0.601	0.54			
6	1.015	0.511	0.05	1.308	0.616	0.04			
7	0.122	0.560	0.83	0.252	0.642	0.70			
PREVIS2	1.119	0.248	<0.001	0.838	0.296	0.005	-		
SZOO	0.549	0.095	<0.001	0.581	0.103	<0.001	0.602	0.119	<0.001
PASTALLC	0.522	0.099	<0.001	0.533	0.106	<0.001	0.551	0.121	<0.001
SZOO*PASTALLC	0.099	0.034	0.004	0.096	0.037	0.009	-0.104	0.040	0.01
MEQUIR	0.768	0.253	0.002	0.795	0.275	0.004	0.948	0.327	0.004
PREVMEQ	-0.699	0.316	0.03	-0.632	0.332	0.057	-		
SERRATIAC	0.648	0.290	0.03	0.604	0.312	0.053	-		
Scaler (MELR only)				0.7007	0.224				

SZOO: *S. zooepidemicus* log₁₀ colony forming units/ml tracheal wash sample (cfu) (0-7). PASTALLC: *Pasteurella/Actinobacillus* spp. log₁₀ cfu (0-7). SERRATIAC: *Serratia* spp. log₁₀ cfu (0-7). MEQUIR: presence of *M. equirhinis* (0,1). PREVMEQ: presence of *M. equirhinis* in the previous time period (0,1).

DISCUSSION

All three methods of analysis demonstrate the close association between disease of the equine lower respiratory tract and infection with *S.zooepidemicus*, *Pasteurella/Actinobacillus* spp. and *M.equirhinis*. Important also was the lack of association between this disease and common equine viruses, including equine herpesviruses 1 and 4 and equine rhinovirus-1. Data on equine influenza are not yet available.

As previously noted for analysis of similar data (Wood and Burrell 1993) and other comparisons (Atwill *et al.*, 1995), results from OLR and MELR were quantitatively and qualitatively similar, although coefficient standard errors were larger in MELR. In contrast, results from CLR were both quantitatively and qualitatively different. However, this is due to great extent to the fact that CLR only uses data from subjects that had episodes of health and disease (Wood & Burrell, 1993). Despite these differences, results from CLR are broadly similar, with the effects of only one infection, *Serratia* spp., not being included in the final model, this being the one with the lowest incidence of the significant variables (1.3% overall).

Interestingly, some significance levels for age and some individual trainers were smaller in MELR compared to OLR. However, trainer is a special variable in terms of the random effects, as horses were nested within training yard. Future approaches to the analysis of these data will include nested random effects models (Hedeker & Gibbons, 1984), treating horse as a random effect nested within training yard. As ages were classified in terms of years, there is a similar relationship here as well.

One criticism of the methods of analysis presented here is that they ignore the correlation of data within each time period. Analysis of this dataset has demonstrated that there is greater intraclass correlation associated with time period than there is associated with horse (data not shown). Although nested MELR models are available and could simultaneously adjust for the effects of clustering within trainer and horse, clustering within time is nested within neither of these variables.

It seems that, if only due to their more efficient use of all the data, MELR models are to be preferred to CLR models. Both are to be preferred to OLR which ignores the clustering of data. It would be of interest to compare results from these analyses with those from GEE.

ACKNOWLEDGEMENTS

These studies were supported by the Veterinary Advisory Committee of the Horserace Betting Levy Board and the Animal Health Trust. They would not have been possible without the assistance of veterinary surgeons M.H.Burrell, D.Dugdale, R.Pilsworth, R.Hopes, M.Shepherd, B.Herinkcx and J.Main. The authors are particularly grateful to racehorse trainers Sir Mark Prescott, L.M.Cumani, S.Dow, A.Scott, E.Dunlop, J.H.M.Gosden, Major W.R.Hern and their staff for their involvement in these studies.

REFERENCES

- Atwill, E.R., Mohammed, H.O., Scarlett, J.M. & McCulloch C.E., (1995). Extending the interpretation and utility of mixed effects logistic regression models. *Preventive Veterinary Medicine* 24, 187-201.
- Burrell, M.H., Whitwell, K.E., Wood, J.L.N. and Mumford, J.A. (1996) Respiratory Disease in Thoroughbred Horses in Training: The Relationship between Disease and Viruses, Bacteria and Environment. *Veterinary Record* 139, 308-313.
- Dean, A.G., Dean, J.A., Coulombier, D., Brendel, K.A., Smith, D.C., Burton, A.H., Dicker, R.C., Sullivan, K., Fagan, R.F. & Amer, T.G. (1994). Epi Info, Version 6: a word processing, database, and statistics program for epidemiology on microcomputers. Centers for Disease Control and Prevention, Atlanta, Georgia, U. S. A.
- EGRET (1989) Statistics and Epidemiology Research Corporation, Seattle, Washington, USA.
- Hedeker, D. & Gibbons, R.D. (1994) A random effects ordinal regression model for multilevel analysis. *Biometrics* 50, 933-944.
- Hosmer, D.W. & Lemeshow, S. (1989). *Applied Logistic Regression*. John Wiley, New York.
- Liang, KY. & Zeger, S.L. (1986) Longitudinal Data Analysis Using Generalised Linear Models. *Biometrika* 73, 13-22.
- Montgomery, J.N., Lehmann, D., Smith, T., Michael, A., Joseph, B., Lupiwa, T., Coakley, C., Spooner, V., Best, B., Riley, I.D. and Alpers, M.P. (1990). Bacterial colonisation of the upper respiratory tract and its association with acute lower respiratory tract in highland children of Papua New Guinea. *Journal of Infectious Diseases* 12, Suppl 8, S1006-S1016.
- Rossdale, P.D., Hopes, R., Wingfield-Digby, N.J. & Offord, K. (1985). Epidemiological study of wastage among racehorses 1982 and 1983. *Veterinary Record* 116, 66-69.
- Wood, J.L.N. & Burrell, M.H. (1993) Epidemiological studies with multiple episodes as outcome: a suggested analytical technique using a cohort study of respiratory disease in horses. *Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine*, p.213-224
- Wood, J.L.N., Burrell, M.H., Roberts, C., Chanter, N. & Shaw, Y. (1993) Streptococci and Pasteurella spp. Associated with Disease of the Equine Lower Respiratory Tract. *Equine veterinary Journal* 25, 314-317
- Wood, J.L.N., Chanter, N., Newton, J.R., Burrell, M.H., Dugdale, D., Windsor, H.M., Windsor, G.D., Rosendal, S. & Townsend, H.G.G.(1997) An outbreak of respiratory disease in racehorses associated with *Mycoplasma felis* infection. *Veterinary Record* 140, 38-192