Use of Markov regression modelling to explore seasonality of colic in the horse

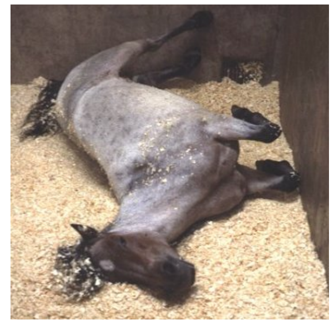
Archer, D.C., Pinchbeck, G.L., Proudman, C.J., Clough, H.E.

Epidemiology Group, Department of Veterinary Science, University of Liverpool, Leahurst, Neston, Wirral .CH64 7TE

THE UNIVERSITY of LIVERPOOL

Introduction:

- The effect of season on the prevalence of colic in the horse is unclear. This may be because statistical methods for describing temporal patterns in data are relatively novel in clinical veterinary research.
- Traditional time-series modelling approaches are based upon Normal distribution assumptions and approximations but in the case of rare diseases, such as horses with specific types of colic, the counts in each month are small making the applicability of such models limited
- Alternative approaches that respect the discrete nature of the count data should be considered in such cases; one suitable method consists of modelling the count data directly through a discrete-valued time series – based approach



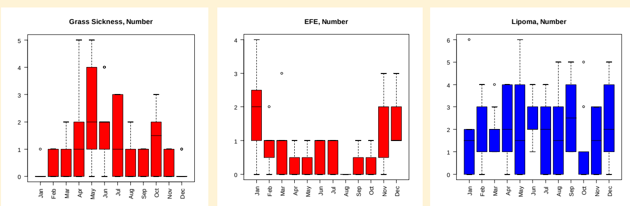
Aim:

- To demonstrate the use of the Markov regression modelling approach of Zeger and Quaquish (1988) to investigate the temporal patterns of 3 types of colic in the horse: Equine Grass Sickness (EGS), Epiploic Foramen Entrapment (EFE) and strangulation of intestine by pedunculated lipomas (PL)
- We hypothesised that EGS and EFE would show a seasonal pattern whereas PL would be a random event with no evidence of a seasonal effect

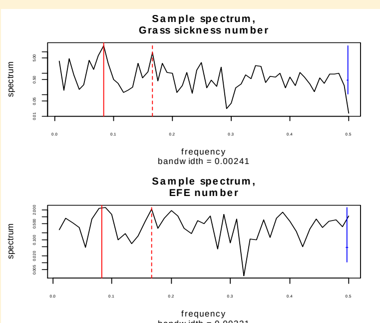
Materials and methods:

Exploratory data analysis

For each colic type, the data were aggregated by year and a box plot of (a) number and (b) proportion of cases by month was made, to look for informal suggestion of seasonality (evidenced by peaks and troughs). Patterns within (a) and (b) were broadly similar and so we have presented (a) for illustration.



- There was a suggestion of a seasonal effect in EGS (peak in May) and EFE (peak in Dec/ Jan), but those for PL showed no discernible effects.



- The sample spectrum highlights cycles in the data (evidenced by a "tall spike"). A spike was seen at the 12-month frequency for EGS prevalence; there was also a smaller spike at the 6 month frequency. For the EFE data, similar spikes, though less pronounced, were evident. No cyclic effects were observed for the PL data.

Statistical modelling

To examine temporal patterns within each separate colic type, we used the Markov regression modelling approach [1] which extends the ideas of generalised linear modelling [2] into a time series context, and is particularly useful in the presence of (a) discrete-valued data and (b) small samples. The general model for the prevalence of colics of type *j* including seasonality, time trend and dependence on previous observations was:

$$y_{j,t} \sim \text{Binomial}(n_{j,t}, p_{j,t})$$

$$\text{logit}(p_{j,t}) = \alpha_j + \beta_j t + \sum_{k \in K_j} \left[\gamma_{1jk} \sin\left(\frac{2\pi t}{k}\right) + \gamma_{2jk} \cos\left(\frac{2\pi t}{k}\right) \right] + \sum_{i=1}^{q} \rho_{j,i} \left(\frac{y_{j,t-i}}{n_{j,t-i}} \right) \quad (1)$$

Here, K_j for colic type j is either 12 or the set {6, 12} to represent 12, or 6- and 12-month cyclicity respectively.

Model fit was assessed via the Bayesian information criterion (BIC) as suggested in [3], and defined, for an arbitrary model M to be:

$$\text{BIC} = -2 \cdot \text{log-likelihood}(M) + npar_M + \log(n_M)$$

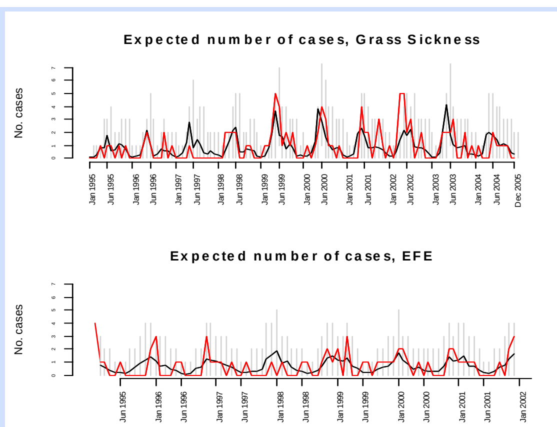
where *npar* is the number of parameters in model, and *n* the number of observations.

Results:

EFE: The lowest BIC came from a model including 12-month seasonal components (peak Dec/Jan) and proportion positive at the previous time point (BIC = 178.25).

Grass sickness: The lowest BIC came from a model including 6- and 12-month seasonal components (BIC = 275.50.) This provides evidence of a primary peak in cases (May) and a secondary peak (September).

PL: No model improved upon a constant mean over time (BIC = 391.18). Either the probability of PL genuinely is constant, or unmeasured covariates may help to explain prevalence.



- For the prevalence model of each colic type, the full model included 6- and 12-month seasonal components, prevalence at the previous time point and time trend.

Key to plots:

data
model fits
exact binomial 95% error bars

Conclusions:

- This particular time series modelling approach respects the discrete nature of the prevalence data and has allowed us to establish formally the existence of a statistically significant seasonal effect in specific types of colic presented at our clinic.
- Use of time-series modelling has confirmed that EGS has a seasonal effect as demonstrated by other workers using different methods of analysis.
- PL showed no seasonal effect consistent with our original hypothesis.
- Knowledge of a significant seasonal effect of EFE has generated hypotheses regarding the possible aetiology and potential risk factors for this condition and these are currently under investigation.

References:

[1] Zeger, S. L. and Qaqish, B. (1988) Markov regression models for time series: a quasi-likelihood approach. Biometrics Vol. 44, pp. 1019-1031.
[2] McCullagh, P. and Nelder, J. A.. (1983) Generalised Linear Models. Chapman and Hall, Monographs on Statistics and Applied Probability.
[3] MacDonald,I. L. and Zucchini, W. (1997) Hidden Markov and other models for discrete-valued time series. Chapman and Hall, Monographs on Statistics and Applied Probability.