Epidémiol. santé anim., 1997, 31-32

# APPLICATION OF FRAILTY MODELS IN MULTIVARIATE SURVIVAL DATA : THE RISK OF CRYPTOSPORIDIUM IN DAIRY CATTLE

Mohammed H.O.[1], Wade S.E.[2]

*L'objectif de cet article est d'examiner l'importance empirique des modèles fractiles dans l'analyse des données longitudinales et d'évaluer leur impact sur l'estimation du risque. Nous avons mené une étude longitudinale pour identifier les facteurs de risque de l'infection des vaches laitières par Cryptosporidium parvum. Ce parasite représente un danger pour la santé humaine. Des échantillons de matière fécale ont été analysés pour rechercher ce parasite en utilisant la méthode de centrifugation. Plusieurs facteurs démographiques, de gestion des élevages, et de santé ont été explorés dans cette étude. Les résultats sont analysés à l'aide du modèle semiparamétrique de risque proportionnel de Cox. L'incorporation de variable latente permet d'estimer le fonction de risque avec une moyenne de distribution finie ou infinie. Ce qui permet d'ajuster les effets estimés sur des variables non observées. Cette méthode a en général un effet sur les estimations sans pour cela entraîner des changements dans le sens des relations.*

## INTRODUCTION

In many longitudinal studies in veterinary epidemiology the sampling of an outcome may be clustered in subgroups, for example, in a sample of related cattle in herds, in matched horses in an intervention study, or in a study with repeated measurements in animals or in flocks. Such data typically have correlated failure times. Such a correlation, if ignored, has the potential of biasing the inferences made from the effect estimates of putative risk factors (Hougaard, 1984). The common approach for the analysis of survival data assumes a homogenous population of study units with every animal in the study population having the same survival function (Cox, 1972). But many sampled populations, especially if sampled in clusters, are not homogenous or correlated. Ignoring this correlation in survival data analysis will result in a bias towards zero in the parameter estimates and in a negative bias in the estimated time dependence.

The expected correlation between responses in longitudinal studies occurs because they are dependent on exogenous factors that are associated with these responses. Conditioning on observed set of these factors by controlling for their effect in the analysis and including them as covariates in a regression analysis will sometimes achieve approximate conditional independence. However, more often that this correlation in the response arises from both observed and unobserved risk factors. The heterogeneity in the response due to the unobserved factors is commonly referred to as frailties (Vaupel et al., 1979; Hougaard 1987). Ignoring this heterogeneity in the analysis of survival data with constant covariates results in a negative bias, towards the null, in the parameters being estimated (Pickles and Croucheley 1995). Therefore, adjusting for such frailty will lead to the increase in the magnitude of the estimates but not the directional relationship.

Several methods have been proposed to address the problem of heterogeneity in data collected in longitudinal studies by incorporating a random effect in the baseline hazard function to adjust for the heterogeneity due to the unobserved factors. The random effect, commonly referred to as frailty, is included in the model as a multiplicative adjustment to the baseline hazard (Hekman and Singer 1984; Haugaards 1989). The standard Cox-proportional hazard model is specified as follows:

$$h_{ij}(t_{ij}) = \lambda_j^0(t_{ij})\mu_{ij}$$

where $\mu_{ij} = \exp(\beta_x X_i + \beta_z Z_{ij})$, and $\lambda_j^0(t_{ij})$ is the baseline hazard for the jth response time of the study unit i. If we

consider the effect of frailty then the above model will be specified as:

$$h_{ij}(t_{ij}) = \lambda_j^0(t_{ij})\exp(\beta_x X_i + \beta_z Z_{ij} + \alpha_{ij})$$

where $\alpha_{ij}$ is the specific frailty.

[1] Section of Epidemiology, Department of Clinical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853

[2] Parasitology Section, Diagnostic Laboratory, College of Veterinary Medicine,Cornell University, Ithaca, New York 14853

The purpose of this manuscript was to examine the emperical importance of the use of frailty models in the analysis of lonitudinal data evalaue the impact of the frailty models on the risk estimates.

## MATERIAL AND METHODS

We carried out a longitudinal study to identify risk factors that predispose calves to infection with *Cryptosporidium parvum* in dairy herds. This parasite is a potential environmental hazard to human health. Baseline data were collected from 40 farms that housed 2,400 animals. The herds were followed for a complete year and fecal samples were collected from calves repeatedly to determine the infection status with *C. parvum*. The fecal samples were analyzed for the presence of the parasite using the quantitative centrifugation flotation method. Data on putative risk factors that were hypothesized to influence the likelihood of infection with *C. parvum* were collected using a questionnaire. Several putative demographic, management, and health factors were considered in the analysis.

The data were analyzed using the semiparametric Cox's proportional hazard model described above. In that model $t_{ij}$ is the time to become infected with *C. parvum*, $\beta_0$ is the baseline hazard, $\beta_i$ is the effect of the hypothesized risk factors X, $\alpha_{ij}$ is the random effect, $e_{ij}$ is the residual. When one accounts for the potential effect of the herd on the animals's risk of infection, the ratio of the hazard for an animal which becomes infected at a given period of time, $\mu_{ij}$ is estimated as follows:

$$\mu_{ij} = \frac{\exp(\beta_0 + \beta_{ij}x_{ij} + \alpha_i)}{\sum \exp(\beta_0 + \beta_{ij}x_{ij} + \alpha_i)}$$

## RESULTS AND DISCUSSIONS

The significance of these factors and their effects on the likelihood of infection with *C. parvum* was initially evaluated using the Cox's proportional hazard model ignoring the potential heterogeneity due to factors other than those observed. The initial analysis was followed by the use of survival models that employ a mixture likelihood to integrate out the frailty effects due to the unobserved covariates. This was achieved by incorporating a random effect component into a baseline survival model to adjust for the population heterogeneity.

The results of the two models, one with the frailty and the one without frailty, generally agreed on the risk factors that were associated with the likelihood with *C parvum*. However, in close examination of the two models, there was evidence for significant variation in both the individual intercept and the linear time trend. In this analysis, the frailty was not just a nuisance but a prime interest in the analysis.

The incorporation of frailty in the multivariate survival analysis allows for the hazard function to be estimated with either finite or infinite mean distribution. Such incorporation will adjust for the effect of unobserved covariates on the effect estimate of the observed covariates. It generally influences the magnitudinal effect of the observed covariates and not the directional relationship to a specific hazard.

## REFERENCES

Cox D.R., 1972. Regression models and life tables. Journal of The Royal Statistical Society, Series B 34, 187-220.

Heckman J. J., Singer B. A. 1984. A method for minimizing the impact of distributional assumptions in econometric models of duration data. Econometrica 52, 271-320.

Hougaard P. ,1986. Survival models for heterogenous populations derived from stable distribution. Biometrika 73, 387-396.

Hougaard P., 1986. A class of multivariate failure ime distributions. Biometrika 73, 671-678.

Pickles A., Crouchley R. 1995. A comparison of frality models for multivariate survival data. Statistics in Medicine 14, 1447-1461.

Vaupel J. W., Manton J. K., Stallard E. 1979. The impact of heterogenity in individual frailty on the dynamics of mortality. Demography 16, 439-454.