Modelling interval-censored, clustered cow udder quarter infection times of dairy cows

Goethals, K.[1], Ampe, B.[1], Berkvens, D.[2], Laevens, H.[1], Janssen, P.[3], and Duchateau, L.[1]
[1]Ghent University; [2]Institute of Tropical Medicine; [3]Hasselt University.

Abstract

A longitudinal study was conducted in dairy cows to study the effect of some parameters (e.g. parity) on time to intramammary infection by different bacteria. Dairy cows were assessed monthly for the presence of these bacteria at udder quarter level. When analysing these time to infection data, the ordinary survival model needs to be adapted in two ways. First, the time to infection is not known exactly; it is only known to have occurred within a certain interval. Second, the four udder quarters are clustered within a cow. We developed a new technique to handle the problem of interval-censoring and clustering simultaneously. The technique is based on the parametric Weibull frailty model introducing the cow in the model as a Gamma distributed random effect or frailty. In this setting it is possible to integrate out the frailties exactly to obtain the marginal likelihood that can then be maximised to obtain parameter estimates. Furthermore, the frailty model also supplies a measure on the clustering of infections within the cow: more clustering can be expected with a more infectious agent.

Introduction

The infectious disease data set considered in this paper, the mastitis data, corresponds to infection times of individual cow udder quarters with a bacterium (Laevens et al., 1997). Obviously, the four udder quarters are clustered within a cow (Adkinson, 1993) and udder quarters are sampled only monthly, generating interval-censored data. We want to investigate the effect of covariates that change within cow (e.g. front and rear udder quarters) and covariates that change between cows (e.g. primiparous versus multiparous). But also the level of clustering expressed by the heterogeneity parameter in the model is of interest because it is amongst others a measure of the infectivity of the agens which causes the disease.

To handle the problem of interdependence for right-censored observations different models have been used among which the frailty model is the standard. For the analysis of independent interval-censored data a number of inferential techniques have been described in the literature (Radke et al., 2003; Collet, 1994). Analysis methods for settings where observations are at the same time correlated and interval-censored received less attention.

In this paper we study an extension of the parametric shared gamma frailty model to interval-censored data. We show that a closed form of the marginal likelihood can be obtained by integrating out the gamma-distributed frailties.

Methods

Independent udder quarter infection times can be modelled by the following proportional hazards model

$$h_{ij}(t) = h_0(t) \exp(\bullet\, {}^t\mathbf{x}_{ij}),$$

with $h_{ij}(t)$ the hazard at time *t* for udder quarter *j* of cow *i*, $h_0(t)$ the baseline hazard at time *t*, $\mathbf{x}_{ij}$ the vector of covariates for the corresponding udder quarter and • the vector of covariate effects. In

Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics, 2006
Available at www.sciquest.org.nz

case that infection times are either exact or right-censored parameters can be obtained by maximising the following likelihood

$$L(\bullet,\bullet,\bullet) = \prod_{i=1}^{n} \prod_{j=1}^{4} \left[h(t_{ij})\right]^{d_{ij}} S(t_{ij}),$$

where $\bullet_{ij}$ is the censoring indicator (0: censored, 1: event) and $S(t_{ij})$ the survival function at time *t* for udder quarter *j* of cow *i*.

But it's clear that udder quarter infection times are not independent, they are correlated within the cluster cow. To handle the problem of interdependence between udder quarters we add a frailty term to the proportional hazards model:

$$h_{ij}(t) = h_0(t)\, u_i \exp(\bullet^{t}\mathbf{x}_{ij}),$$

with u_i the effect of cow *i* . The frailties $u_1, \ldots, u_N$ are independent realizations from a one parameter gamma density with mean one and variance •:

$$f_U(u_i) = \frac{w_i^{1/q-1}\exp(-w_i/q)}{q^{1/q}\,\Gamma(1/q)}.$$

Furthermore it's obvious that herds aren't visited every day. Thus interval censored data are generated. In case data are right or interval-censored the likelihood takes the following form:

$$L(\bullet,\bullet,\bullet) = \prod_{i=1}^{n} \prod_{j=1}^{4} \left[S(L_{ij}) - S(U_{ij})\right]^{d_{ij}} \left[S(R_{ij})\right]^{(1-d_{ij})},$$

where L_{ij} and U_{ij} denote the lower and upper bound of the interval and R_{ij} denotes the right-censoring time.

To handle the problem of correlated interval-censored infection time data we combine these two extensions of the proportional hazards model for independent exact and right-censored data into a single model with marginal likelihood

$$L(\bullet,\bullet,\bullet) = \prod_{i=1}^{n} \int_{0}^{\infty} \prod_{j=1}^{4} \left[S(L_{ij}) - S(U_{ij})\right]^{d_{ij}} \left[S(R_{ij})\right]^{(1-d_{ij})} f_U(u_i)\, du_i.$$

Throughout this paper we assume a Weibull distribution for the hazard, but the method can easily be extended to other parametric forms of the baseline hazard. Assuming a gamma distribution for the frailties enables us to integrate out the frailties analytically and therefore the marginal likelihood can be maximised to obtain parameter estimates. As the second partial derivatives can be obtained for all parameters, an explicit expression for the information matrix is available, from which an estimate of the asymptotic variance-covariance matrix can be obtained.

Results

To illustrate the technique a sample of 100 cows randomly selected from a larger dataset is analysed. First the effect of the covariate parity on the time to infection with a bacterium was investigated. We compared our technique with two commonly used methods to handle the problem of interval-censoring: imputation of midpoint and upper bound (Table 1).

Table 1 Parameter estimates and their standard errors with parity as covariate

	• (SE)	• (SE)	• (SE)	• (SE)
Exact	1.7718 (0.3031)	0.8108 (0.2087)	1.9239 (0.1099)	0.3139 (0.3299)
MP	1.7927 (0.3043)	0.8379 (0.2145)	1.9791 (0.1064)	0.3167 (0.3284)
UB	1.8921 (0.3126)	0.5441 (0.1418)	2.4036 (0.1300)	0.2463 (0.3339)

We also investigated whether there was a difference in time to infection between front or rear udder quarters (Table 2).

Table 2 Parameter estimates and their standard errors with front or rear udder quarter as covariate

	• (SE)	• (SE)	• (SE)	• (SE)
Exact	1.7836 (0.3033)	0.8956 (0.1653)	1.9287 (0.1098)	0.1798 (0.1219)
MP	1.7981 (0.3041)	0.9264 (0.1713)	1.9807 (0.1060)	0.1774 (0.1194)
UB	1.8999 (0.3125)	0.5753 (0.1042)	2.4080 (0.1297)	0.1848 (0.1195)

In both cases the parameter estimate of • is above 1, and the hazard is thus increasing with time. The hazard ratio of multiparous cows versus primiparous cows is exp(0.3139)=1.37; the 95% confidence interval is [0.72;1.88]. The hazard ratio of front versus rear udder quarters is exp(0.1798)=1.20; the 95% confidence interval [0.94;1.52]. The estimate for • is 1.8; the corresponding estimate for Kendalls tau (Kendall, 1938) is • / (• +2)=0.4736. Thus, infection times within the cow are highly correlated, as could be expected.

It might seem that the proposed method has no advantage over imputation of midpoint. Simulation studies however show that in case of a decreasing hazard imputation of midpoint fails completely giving biased estimates. The proposed method outperforms imputation of upperbound.

Discussion

We proposed a shared gamma frailty model to handle correlated and interval-censored time to infection data. Assuming a gamma distribution for the frailty enables us to integrate out the frailties analytically and to obtain a closed form expression for the marginal likelihood, which can then be maximised to obtain parameter estimates. Simulation studies show that the proposed technique outperforms the imputation techniques.

References

Adkinson, R.W., Ingawa, K.H., Blouin, D.C. and Nickerson, S.C. (1993). Distribution of clinical mastitis among quarters of the bovine udder. *Journal of Dairy Science*, 76, 3453-3459.

Collet, D. (1994). *Modelling survival data in medical research*. London: Chapman & Hall.

Kendall, M.G. (1938). A new measure of rank correlation. *Biometrika*, 30, 81-93.

Laevens, H., Deluyker, H., Schukken, Y.H., De Meulemeester, L., Vandermeersch, R., De Meulenaere, E. and De Kruif, A. (1997). Influence of parity and stage of lactation on the somatic cell count in bacteriologically negative dairy cows. *Journal of Dairy Science*, 80, 3219-3226.

Radke, B.R. (2003). A demonstration of interval-censored survival analysis. *Preventive Veterinary Medicine*, 59, 241-256.