Beyond Cox: 1. Hazard functions and time varying effects in parametric survival models

I. R. Dohoo*, H. Stryhn, J. Sanchez
Department of Health Management, University of PEI, Charlottetown, PEI, C1A 4P3.

Summary
While parametric models for survival data offer some advantages over semi-parametric models (Cox proportional hazards model), it is important that the distribution of the baseline hazard be specified correctly. This paper summarises methods for evaluating the shape of the baseline hazard and discusses the potential effect of frailties (random effects) on the observed distribution. In addition, the effects of predictors may vary over time and one approach to assessing the nature of those changes (Aalen's linear hazards model) is presented.

Introduction

The Cox proportional hazards model is the most widely used approach for analysing time to event data because it requires no assumption about the distribution of the baseline hazard. However, it achieves this flexibility at a price. Since a Cox model is based only on the rank ordering of failures in a dataset, rather than the exact time of failure, it does not utilize all of the data available. This limitation is particularly evident if any of the predictors in the model are time-varying and may experience several changes between times that failures occur. Parametric survival models take into account the exact time of failure and censoring events and all changes in predictors between those events but are only valid if the distributional assumptions about the baseline hazard are fully met. Failure to ensure that the baseline hazard has the appropriate form for the distribution assumed in a parametric model may lead to seriously biassed estimates of coefficients in the model.

Parametric survival models are based on an assumed function of the baseline hazard with the more commonly used functions being ones which gives rise to exponential, Weibull, log-normal, log-logistic and gamma distributions of survival times. Some of these distributions (eg. exponential, Weibull) give rise to proportional hazards models and can be written as:

$$h(t) = h_0(t)e^{\beta X}$$

where the baseline hazard (h_0) has the functional form required to give rise to the assumed distribution of survival times (eg. constant hazard for exponential survival times). Exponentiated coefficients from these models are "hazard ratios" (hr) that represent the multiplicative factor by which the hazard of event occurrence is increased or decreased by the predictor. Not all parametric models are proportional hazards models, so they are sometimes presented as accelerated failure time models in which the predictors act multiplicatively on the time to failure:

$$t = e^{\beta X}\tau$$

where τ has the specified distribution of survival times.

Evaluating the shape of the baseline hazard

In some cases, you may be reasonably sure of the shape of the baseline hazard based on your biological knowledge of the system. In other cases, you will want to base your assumption about the shape on evidence from the observed data. Fig 1 (dotted line) shows a log-normal hazard function that was used to simulate data (n=2000) with a single dichotomous predictor (hr=2). One simple approach to estimating the shape of the baseline hazard is to generate a smoothed curve of a Kaplan-Meier estimate of the hazard for those observations with all covariates equal to zero (Fig. 1 dashed line). This approach is limited by the availability of data and consequently, the graph shown has been limited to survival times <50 days.

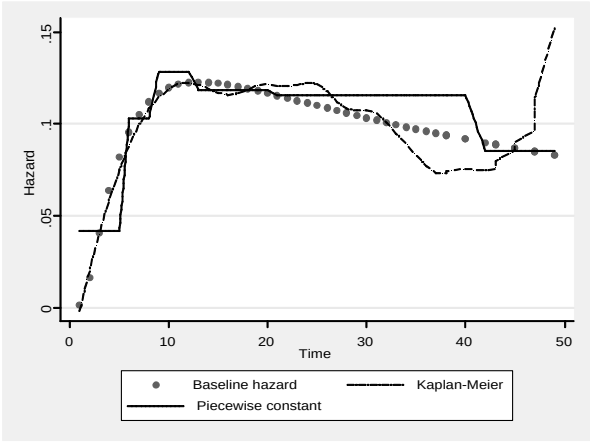

Fig. 1 Kaplan-Meier and piecewise-constant estimates of a log-normal baseline hazard

An alternative approach is to divide the survival times up into intervals, and assume that the hazard is constant within each of those intervals. To do this, the data were divided into intervals at 5, 8, 12, 20 and 40 days (containing 18%, 29%, 26%, 19%, 7% and 1% of the failure times respectively) and an exponential model fit with indicator variables for the time intervals. The estimated baseline hazard in each interval is shown in Fig. 1 (solid line). This approach is very flexible but requires a reasonable sized sample in each of the time intervals.

Frailty (random) effects can have a profound influence on the observed baseline hazard (Cleves et al, 2002; Gutierrez 2002). Frailties are random effects that are assumed, most commonly, to act multiplicatively (additive frailty models are also possible) on the survival times. Frailties may represent the effect of unmeasured covariates on survival times for individuals in the population. Frailties can be assumed to be unique for each individual or common to a group of individuals (eg. cows within a herd). Frailty models can be written as:

$$h(t)_i = h_0(t)e^{\beta X}\alpha_i \qquad \text{or} \qquad t_i = e^{\beta X}\tau\alpha_i$$

where α is the frailty effect and i designates the individual in an "individual frailty" model or the group designator in a "shared frailty" model. Commonly used distributions for frailties are gamma, inverse-Gaussian and log-normal. When a frailty effect is added to a model, the coefficients for the predictors represent their effect at time = 0 but the effects of predictors will diminish over time.

Individual frailties may have a pronounced effect on the observed hazard. Data were simulated with a constant hazard and a gamma frailty. Fig 2 shows the estimates of

the baseline hazards using the Kaplan-Meier and piecewise constant approaches described above. Although the underlying hazard was constant, it appears to decrease with time since, as more "frail" individuals fail early and the population becomes increasingly robust as time moves on.

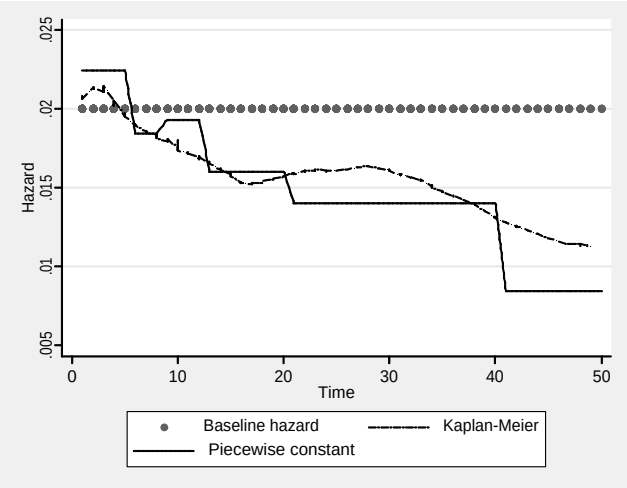

Fig. 2. Effect of frailty on observed hazard function

Effect of predictors over time

In all survival models, the nature of the effect of predictors over time must be considered. In proportional hazards models, the effects are assumed to be constant, although this assumption can be relaxed by incorporating interaction terms between predictors and some function of time. An Aalen linear hazards model (Hosmer & Royston, 2002) can be used to evaluate how the effects of predictors vary with time and consequently how predictor-time interaction terms should be constructed. This procedure produces a plot of the cumulative regression coefficient for a predictor over time. Portions of the graph where the line is flat indicate time periods in which the predictor had no effect. Points where it is increasing or decreasing linearly indicate constant effects over that period. Fig 3 shows the cumulative regression coefficient for treatment effect from data derived from a clinical trial of the effects of prostaglandin administered at the start of the breeding period in dairy cows. As can be seen, treatment had a pronounced effect about day 3, followed by fewer conceptions in the treated group up to about day 25, another positive effect from days 25 to 33 and then relatively little effect (flat line).

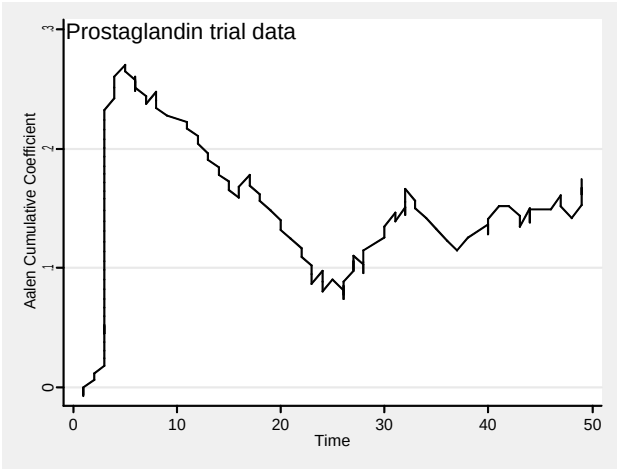

Fig. 3. Effect of prostaglandin treatment on conception in dairy cows.

References:

1. Cleves,M.A., Gould,W.W., Gutierrez, R.G. 2002. An Introduction to Survival Analysis Using Stata. Stata Press, College Station, TX.
2. Gutierrez.R.G. 2002. Parametric frailty and shared frailty survival models. Stata J. 2, 22-44.
3. Hosmer.D.W., Royston,P. 2002. Using Aalen's linear hazards model to investigate time-varying effects in the proportional hazards regression model. Stata J. 2, 331-350.