Joint modeling of a survival and a count response

A.A. Sunethra and M.R. Soriyarachchi
Department of Statistics, University of Colombo.

Survival and incidence are two common response variables in medical data. Thus, statistical models with responses of survival and incidence (count) are common in medical data analysis, though these two responses have not been considered in the literature as a bi-variate response within a single model. However, in many cases these two responses can be correlated, *i.e.* survival time of a patient can have some bearing with the rate of incidence of the disease. For example, diseases that occur rarely can have a shorter survival time or vice versa. When two responses are correlated, joint modelling of them simultaneously within a single model would provide improved results since such models take into account the correlation between the two responses. This study considered formulating a methodology for jointly modelling a survival and frequency (count) response. The difficulty of obtaining a joint distribution between the two variables for survival and Poisson, due to the former being continuous with censored observations and the latter being discrete was overcome by survival times being fitted as a Poisson random variable which was realistic due to the equivalence of the log-likelihoods of survival times and Poisson random variables under the assumption of proportional hazards. This required specifying the censoring indicator of the survival time as a Poisson random variable. Then, the joint density of the two Poisson distributions was considered as 'bi-variate Poisson' for fitting the joint model. Using R software, a bi-variate Poisson model was fitted for a partially simulated data set, which consisted of actual survival times of some leukemia patients with two treatment groups and a positively correlated count variable was simulated. It was assumed survival times are exponentially distributed. Model fitting revealed that only significant predictor was 'treatment' and the joint model was better than two univariate models with respect to the BICs of the models. The predicted correlation between the two responses was quite closer to the actual correlation of the data. Using different parametric distributions and semi-parametric models for survival times and using different distributions for the joint distribution of two Poisson random variables can be suggested as further developments.