A new bivariate survival model including a non-susceptible fraction

Wienke A[1], Ripatti S[2], Palmgren J[2], Yashin AI[3]
[1]*Institut für Medizinische Epidemiologie, Biometrie und Informatik, Universität Halle, Deutschland*
[2]*Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden*
[3]*Center for Demographic Studies, Duke University, Durham, USA*
andreas.wienke@medizin.uni-halle.de

Traditional analysis of time-to-event data is based on two main assumptions: first, independence of observed event times and second, all individuals are susceptible to the event of interest and will eventually experience this event if the follow-up is sufficiently long.
Various approaches are investigated to overcome the first restriction of independence, known as multivariate survival models. Available statistical models fall into two broad classes - marginal and frailty models. Marginal methods of analysis specify models for the effect of covariates on the hazards of the individual events (the margins), taking into account the fact that observed event times are correlated but without the need for explicitly modelling this correlation. The association between the events is considered as a nuisance parameter. As with the analysis of longitudinal data, regression parameters are estimated from generalized estimating equations, and the corresponding variance-covariance estimators are corrected properly to account for the dependence structure.
The marginal approach is ideal for making inference on the population average effect of risk factors on failure time. However, it provides no insight into the multivariate relationship among failure times. These types of questions are answered by frailty models, considering the association between various events explicitly. In general, frailty models have an intuitive appeal and provide insight into the relationship between failures, and we will focus on this approach in the present paper.
Frailty models have been used frequently for modelling dependence in multivariate time-to-event data. The dependence usually arises because individuals in the same group (family, litter, study centre) are related to each other or because of the multiple recurrence of an event for the same individual. The traditional proportional hazards model can not be applied to these cases. A possible solution to this problem is the use of conditional proportional hazards given the frailty. Here, the variability of lifetimes is formulated as arising from two different sources: first, natural variability, which is included in the baseline hazard function; second, frailty. Lifetimes are conditionally independent given the frailty (as individual random effect), and the frailty term represents unobserved covariates. We assume that, given unobserved frailty, the hazard for each individual survival time follows a proportional hazards model, with the frailty variable (the random effect) acting multiplicatively on the baseline hazard function.
To overcome the second strong assumption in general survival models mentioned above cure models are introduced into survival analysis. Cure models allow for a fraction of individuals in the population who are not susceptible to the event under study. Univariate cure models are well established in the literature, but only a few papers exist dealing with bivariate cure models, combining the concepts of frailty and cure modelling [1-3]. Chatterjee and Shih used a shared frailty model with an extra latent variable to model a non-susceptible fraction. Extending this approach Wienke et al. applied a correlated gamma frailty model. Moger and Aalen suggest a compound Poisson frailty model with random scale, without additional latent variables. But their survival function needs a recursive representation. The main aim of the present paper is to establish a simple bivariate frailty model with cure fraction.
To avoid the limitations of the often used shared frailty models, correlated frailty models are being developed for the analysis of multivariate failure time data [2, 4-6], in which associated random variables are used to characterize the frailty effect for each cluster. Correlated frailty models provide not only variance parameters of the frailties as in shared frailty models, but they also contain additional parameters for modelling the correlation between frailties in each group.
Data on Swedish twins was collected by the Swedish Twin Registry, the largest population based twin registry in the world. Information about onset of cancer was obtained from the National Cancer Registry by regularly matching both registries. Unfortunately, such kind of registers does not commonly contain detailed information about individual risk factors. To account for this missing information in the analysis, frailty models can be used. In these models the frailty variable represents varying levels of risk of different individuals in the study population.
A new frailty model is suggested for analysis of bivariate time-to-event data. The model is an extension of the correlated PVF frailty model (correlated three-parameter frailty model) introduced by Yashin and Iachine [7] and has never before been considered in bivariate survival studies. It is based on a bivariate extension of the compound Poisson frailty model introduced to univariate survival analysis by Aalen [8]. It is also related to the extended correlated gamma frailty model by Wienke [2]. It allows for a non-susceptible fraction in the population, overcoming the common assumption in survival analysis that all individuals are susceptible to the event under study. The survival function of the correlated compound Poisson frailty model is given by:

$$S(t_1,t_2) = S(t_1)^{1-\rho} S(t_2)^{1-\rho} e^{-\frac{\rho(1-\gamma)}{\gamma\sigma^2}\left(1-\left(\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_1))\right)^{1/\gamma}+\left(1-\frac{\gamma\sigma^2}{1-\gamma}\ln(S(t_2))\right)^{1/\gamma}-1\right)^{\gamma}\right)}$$

Parameter γ divides the class of distributions in two major subfamilies: For $\gamma \geq 0$ the distribution is a power variance function distribution (PVF). The extension to $\gamma < 0$ in the univariate case was suggested by Aalen [8] and shown to yield the compound Poisson distribution. The two subclasses are separated by the gamma distribution $\gamma = 0$. The model is applied to breast cancer data of 5857 Swedish twin pairs to estimate the correlations between frailties of both monozygotic (ρ_{MZ}) and dizygotic (ρ_{DZ}) twins as well as the size of the susceptible fraction. Results of the analysis are given in the following table:

	gamma frailty	inverse Gaussian frailty	compound Poisson frailty
γ	0	0.5	-0.05 (0.10)
σ	7.61 (0.47)	7.41 (1.08)	7.03 (0.99)
ρ_{MZ}	0.12 (0.04)	0.20 (0.06)	0.12 (0.04)
ρ_{DZ}	0.10 (0.03)	0.16 (0.05)	0.10 (0.03)
φ	**1.00**	**1.00**	**0.34 (0.40)**
log-L	-5218.64	-5265.76	-5218.46

The inverse Gaussian model is given by γ=0.5. Parameter values of $\gamma < 0$ imply the existence of a non-susceptible fraction in the population. Parameter φ describes the size of the susceptible fraction and σ is the standard deviation of the frailty. The log-likelihood is given in the last row of the table. The Results are compared to the analysis by Wienke et al. [2].

Literatur

[1] Chatterjee N, Shih J. A bivariate cure-mixture approach for modeling familial association in diseases. Biometrics 2001; 57: 779 – 786.
[2] Wienke A, Lichtenstein P, Yashin AI. A bivariate frailty model with a cure fraction for modeling familial correlations in diseases. Biometrics 2003; 59: 1178 – 1183.
[3] Moger TA, Aalen OO. A distribution for multivariate frailty based on the compound Poisson distribution with random scale. Lifetime Data Analysis 2005; 11: 41 - 59.

[4] Yashin AI, Vaupel JW, Iachine IA Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. Mathematical Population Studies 1995; 5: 145 – 159.
[5] Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. Biometrics 2000; 56: 1016 – 1022.
[6] Ripatti S, Larsen K, Palmgren J. Maximum likelihood inference for multivariate frailty models using an automated Monte Carlo EM algorithm. Lifetime Data Analysis 2002; 8: 349–360.
[7] Yashin AI, Begun A, Iachine IA. Genetic factors in susceptibility to death: a comparative analysis of bivariate survival models. Journal of Epidemiology and Biostatistics 1999; 4: 53-60.
[8] Aalen OO. Modelling heterogeneity in survival analysis by the compound Poisson distribution. Annals of Applied Probability 1992; 4: 951 - 972