Use of semiparametric frailty model in analysis of survival data in twins

Annah M. Muli[1], Arief Gusnanto[1], and Jeanine J Houwing-Duistermaat[1,2,3].

[1]Department of Statistics, University of Leeds, United Kingdom.

[2] Alan Turing Institute, United Kingdom

[3] Department of Biostatistics and research support, Utrecht University Medical Center, The Netherlands.

E-mail for correspondence: staammu@leeds.ac.uk

Abstract:

Family based studies help to investigate traits that segregate within families e.g., human longevity which is known to cluster within families. This is attributed to the fact that there is correlation between family members (e.g., twin pairs) as they share genetic and environmental factors. Analysis of such data is challenging due to censoring and correlation among the survival times. The shared frailty model is commonly used for analysis of such correlated survival data. A parametric frailty distribution is assumed, for example the gamma distribution which is computationally convenient. Via simulation we have shown that if the frailty distribution is not correctly specified the estimates of the regression coefficient and survival probabilities may be biased.

We consider a nonparametric specification of the baseline hazard by making use of splines. Simulations showed that replacing the parametric baseline hazard by a flexible baseline hazard can adjust for the incorrect frailty distribution and may improve the estimators of the population survival probability.

We therefore propose to use a semiparametric frailty model to estimate individual specific probabilities of fracture in the next time period given covariates using the TwinsUK data. The event of interest is time to fracture for twins aged 50 years and above. About 1500 people developed a fracture. Results of this analysis will be shown.

Key words: frailty; twins; fractures; survival; misspecification