Models for Survival in Matched Pairs

Andersen, Per Kragh
Department of Biostatistics, University of Copenhagen
Oester Farimagsgade 5
DK-1014 Copenhagen K, Denmark
E-mail: pka@biostat.ku.dk

Gerster, Mette
Department of Biostatistics, University of Southern Denmark
Campusvej 55
DK-5230 Odense M, Denmark
E-mail: mgerster@health.sdu.dk

Madsen, Mia
Unit of Epidemiology, University of Southern Denmark
Campusvej 55
DK-5230 Odense M, Denmark
E-mail: mimadsen@health.sdu.dk

INTRODUCTION

Let T_{i1}, T_{i2} be (possibly censored) event times for the two members of a matched pair, e.g. a pair of twins. Inference for these event times needs to address the possible within-pair association and a possible model with this property is the shared frailty model

$$\lambda_{ij}(t) = Z_i \lambda_0(t) \exp(\beta^T X_{ij}(t)), i = 1, \ldots, n; j = 1, 2.$$

Here, $\lambda_{ij}(t)$ is the hazard function for subject (i, j), $\lambda_0(t)$ the baseline hazard common to all individuals, and $X_{ij}(t)$ are observed covariates for subject (i, j). Finally, Z_i is an unobserved random effect ("the frailty"), which is shared between the two members of the pair. The random frailty creates the desired within-pair association and it is assumed that $(T_{i1} \perp T_{i2}) \mid Z_i$.

Standard inference for this model requires independence between Z_i and $X_{ij}(t)$. We study how violations of this assumption affects inference for the regression coefficients, β, and conclude that substantial bias may occur.

We propose an alternative way of making inference for β by using a fixed-effects models for survival in matched pairs. In this model,

$$\lambda_{ij}(t) = \lambda_{0i}(t) \exp(\beta^T X_{ij}(t))$$

and each pair has its own baseline hazard, $\lambda_{0i}(t)$ which is eliminated from estimating equations via a partial likelihood. Fitting this model to data generated from the frailty model provides consistent and asymptotically normal estimates for β even when $X_{ij}(\cdot)$ and Z_i are dependent.

The methods are exemplified by studying how educational status affects the incidence of cancer in Danish twins.

Notation and Data Generating Model

Define $N_{ij}(t) = I(T_{ij} \wedge U_{ij} \leq t, T_{ij} \geq U_{ij})$ where U_{ij} is a censoring time assumed independent of T_{ij}. Furthermore, define the at-risk indicator $Y_{ij}(t) = I(T_{ij} \wedge U_{ij} \geq t)$ and the filtration $\mathcal{H}_t = \bigvee_{i=1}^{n} \mathcal{H}_t^i$

where

$$\mathcal{H}_t^i = \sigma\{N_{ij}(s), Y_{ij}(s), X_{ij}(s), Z_i \mid j = 1, 2; 0 \leq s \leq t\},$$

i.e. $\mathcal{H}_t$ is the filtration in which the unobserved $Z_1, \ldots, Z_n$ are "known". Correspondingly, let $\mathcal{F}_t$ be the *observed* filtration, i.e.

$$\mathcal{F}_t = \bigvee_{i=1}^{n} \mathcal{F}_t^i = \bigvee_{i=1}^{n} \sigma\{N_{ij}(s), Y_{ij}(s), X_{ij}(s) \mid j = 1, 2; 0 \leq s \leq t\}.$$

Assume that the $\mathcal{H}_t$-intensity for $N_{ij}(t)$ is

(1) $\lambda_{ij}(t) = Z_i Y_{ij}(t) \lambda_0(t) \exp(\beta_0^T X_{ij}(t))$

where $Z_1, \ldots, Z_n$ are independent random variables parametrized by some real-valued parameter θ. Since $\mathcal{F}_t^i \subseteq \mathcal{H}_t^i$, the $\mathcal{F}_t$-intensity for $N_{ij}(t)$ can be obtained via the Innovation Theorem (Andersen et al. 1993, p.80) to be

$$\lambda_{ij}^{\mathcal{O}}(t) = E[Z_i \mid \mathcal{F}_{t-}] \cdot Y_{ij}(t) \lambda_0(t) \exp(\beta_0^T X_{ij}(t)).$$

Note that $E[Z_i \mid \mathcal{F}_{t-}]$ depends on β, $\lambda_0(t)$ and θ. For example, in the case where $Z_i \sim \Gamma(\theta, \frac{1}{\theta})$,

$$E[Z_i \mid \mathcal{F}_{t-}] = \frac{\theta + N_{i\cdot}(t-)}{\theta + \sum_{j=1}^{2} \int_0^{t-} Y_{ij}(s) \lambda_0(s) \exp(\beta_0^T X_{ij}(s)) ds}.$$

The Stratified Cox Model

In the stratified (or fixed-effects) Cox Model it is assumed that the $\mathcal{F}_t$-intensity of the counting process is not $\lambda_{ij}^{\mathcal{O}}(t)$ but

(2) $\lambda_{ij}^*(t) = Y_{ij}(t) \lambda_{0i}(t) \exp(\beta_0^T X_{ij}(t)),$

cf. Holt and Prentice (1974). We claim that fitting this model to the data generated by the above described model will give rise to unbiased estimates of β_0, no matter whether the assumption of independence between X and Z is met. This is due to the fact that the score equation arising from the stratified Cox model will give rise to an unbiased estimating equation for data that are generated from the frailty model as we will demonstrate in the following:
The log-partial Cox likelihood corresponding to (2) is

(3) $$C_\tau(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{2} \int_0^\tau \left[\log \left(\frac{\exp(\beta^T X_{ij}(t))}{\sum_{j=1}^{2} Y_{ij}(t) \exp(\beta^T X_{ij}(t))} \right) \right] dN_{ij}(t)$$

which gives rise to the score equation

(4) $$U_\tau(\beta) = \sum_{i=1}^{n} \sum_{j=1}^{2} \int_0^\tau \left[X_{ij}(s) - \frac{S_{1i}(s, \beta)}{S_{0i}(s, \beta)} \right] dN_{ij}(s) = 0$$

where

$$S_{0i}(t, \beta) = \sum_{j=1}^{2} \exp(\beta^T X_{ij}(t)) Y_{ij}(t)$$

$$S_{1i}(t, \beta) = \sum_{j=1}^{2} X_{ij}(t) \exp(\beta^T X_{ij}(t)) Y_{ij}(t),$$

see e.g. Andersen et al. (1993, Example VII.2.13). Now, define the $\mathcal{F}_t$-martingale $M_{ij}^{\mathcal{O}}(t) = N_{ij}(t) - \Lambda_{ij}^{\mathcal{O}}(t)$ where $\Lambda_{ij}^{\mathcal{O}}(t) = \int_0^t \lambda_{ij}^{\mathcal{O}}(s)ds$. The score function (4) can then be re-written as

$$(5) \quad U_\tau(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{2}\int_0^\tau \left[X_{ij}(s) - \frac{S_{1i}(s,\beta)}{S_{0i}(s,\beta)}\right] dM_{ij}^{\mathcal{O}}(s)$$

$$(6) \quad + \int_0^\tau \lambda_0(s)\sum_{i=1}^{n}\left\{E[Z_i \mid \mathcal{F}_{t-}]\sum_{j=1}^{2}\left[S_{1i}(s,\beta_0) - \frac{S_{1i}(s,\beta)}{S_{0i}(s,\beta)}\cdot S_{0i}(s,\beta_0)\right]\right\} ds,$$

which implies that

$$U_\tau(\beta_0) = \sum_{i=1}^{n}\sum_{j=1}^{2}\int_0^\tau \left[X_{ij}(s) - \frac{S_{1i}(s,\beta_0)}{S_{0i}(s,\beta_0)}\right] dM_{ij}^{\mathcal{O}}(s),$$

in particular that $EU_\tau(\beta_0) = 0$ since $M_{ij}^{\mathcal{O}}(t)$ is an $\mathcal{F}_t$-martingale. Hence, the score equation in the stratified Cox model gives rise to an unbiased estimate of β_0 for data generated by the previously described intensity, (1).

Define

$$S_{2i}(t,\beta_0) = \sum_{j=1}^{2} Y_{ij}(t)\mathbf{X}_{ij}(t)^{\otimes 2}\exp(\beta^T\mathbf{X}_{ij}(t)),$$

$$V_i(t,\beta) = \frac{S_{2i}(s,\beta)}{S_{0i}(s,\beta)} - \left(\frac{S_{1i}(s,\beta)}{S_{0i}(s,\beta)}\right)^{\otimes 2}.$$

Then the matrix of second order derivatives of the log-partial likelihood, $C_\tau(\beta)$, can be written as $-\mathcal{I}(\beta)$ where

$$\mathcal{I}(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{2}\int_0^\tau V_i(s,\beta)dN_{ij}(s)$$

Using the martingale central limit theorem, regularity conditions can be derived under which the following holds for $n \to \infty$:

- The probability that the equation $U_\tau(\beta) = 0$ has a unique solution $\hat\beta$ tends to 1
- $\hat\beta \xrightarrow{P} \beta_0$
- $n^{-1/2}(\hat\beta - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\tau^{-1})$
- $\sup_{t\in[0,\tau]} \|n^{-1}\mathcal{I}(\hat\beta) - \Sigma_\tau\| \xrightarrow{P} 0.$

Here, Σ_τ is defined as the limit in probability of $n^{-1}\langle U(\beta)\rangle(\tau)$ (see below).
The following is a sketch of the proof:
First, note that $U_\tau(\beta_0)$ is an $\mathcal{F}_t$-martingale and that the predictable variation process for $U_\tau(\beta_0)$ can be written as

$$(7) \quad \langle U(\beta_0)\rangle(\tau) = \sum_{i=1}^{n}\int_0^\tau \lambda_0(s)E[Z_i \mid \mathcal{F}_{t-}]\cdot S_{0i}(s,\beta_0)V_i(s,\beta_0)ds.$$

Now, using the Martingale CLT (Andersen et al., 1993 Theorem II.5.1), it can be shown that $n^{-1/2}U(\beta_0)$ converges in distribution to a normal distribution with mean zero and covariance matrix, Σ_τ. So we get (by means of a Taylor expansion argument) that

$$0 = U(\hat\beta) \approx U(\beta_0) - I(\beta_0)(\hat\beta - \beta_0),$$

i.e.

$$\sqrt{n}(\hat{\beta} - \beta_0) \approx \left(\frac{1}{n} I(\beta_0)\right)^{-1} (\sqrt{n})^{-1} U(\beta_0) \approx \Sigma^{-1} (\sqrt{n})^{-1} U(\beta_0)$$

and therefore, $\sqrt{n}(\hat{\beta} - \beta_0)$ is asymptotically normal with mean zero and covariance matrix Σ^{-1}.

Application: Education and breast cancer

We examine the effect of educational attainment on breast cancer by using both a frailty model and a stratified Cox model. The data comprised 16.310 Danish female twins, 6268 monozogotic and 10042 dizogotic. Information was obtained from the Danish Twin Registry, administrative registers in Statistics Denmark, and the Danish Cancer Registry. The population was followed from 1980-2006 (Madsen et al., 2011).

Educational status was dichotomized into primary and secondary/tertiary education, and breast cancer was defined according to ICD-10 (C50) and ICD-7 (170). During the follow-up period, 518 events occurred and 19% of MZ and 27% of DZ twins were discordant on educational status. That amounts to a total of 234 informative observations in the within-pair analysis. Unpaired and within-pair analyses were compared in order to identify potential familial confounding. This is based on the argument that if an unpaired analysis gives rise to an educational difference in the hazard of breast cancer this might be due to unobserved confounding, for example as illustrated in Figure 1. When performing an intrapair analysis, the estimate of educational status is *within pairs*, and hence

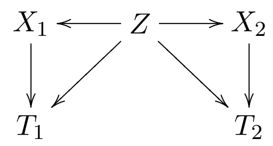

Figure 1: DAG illustrating the confounding problem in the case of twin data. In the present example, X_j denotes the educational status for twin j and T_j the waiting time to breast cancer

unobserved confounding is implicitly controlled for. However, as we have argued above, only in the case where X and Z are independent, can the intrapair comparison be carried out by means of a frailty model. Otherwise, the stratified Cox model should be used.

Consistent with the existing literature on social inequality in cancer, results from unpaired analyses of both models showed an increased risk of breast cancer associated with a high education (HR=1.43 (CI-95% 1.03-1.97) for MZ twins. In contrast, the within-pair results. The Stratified Cox model displayed a massive attenuation of effect in the within-pair analysis: HR=0.82 (95%-CI 0.41-1.67)). For the frailty model, the results of the within-pair analysis were essentially identical to those of the unpaired analysis (HR=1.44 (CI-95% 1.06-1.96)).

Simulations

We generate data from the model

$$\lambda_{ij}(t \mid Z_i) = Z_i \cdot \lambda_0(t) \cdot \exp(\beta \cdot X_{ij}), \quad i = 1, \ldots, n; j = 1, 2.$$

The Z_i are drawn from a Γ-distribution and a log-normal distribution, respectively. In either case the mean and the variance are set to 1. We generate two different scenarios for the covariate: in the first case $X_{ij} \in \{0, 1\}$ is drawn from a Bernoulli distribution with probability $p = 0.5$ (and hence is

			Frailty Model				***Stratified Model***			
n	β	Dist	Bias	SD1	SD2	NA	Bias	SD1	SD2	NA
300	1.2	Gamma	0.005	0.142	0.141	0.001	0.006	0.222	0.218	0.000
75	1.2	Gamma	-0.006	0.283	0.276	0.011	0.010	0.459	0.456	0.000
300	1.2	Log-Normal	0.002	0.127	0.128	0.000	0.011	0.215	0.220	0.000
75	1.2	Log-Normal	0.006	0.257	0.258	0.000	0.007	0.445	0.453	0.000

Table 1: Simulation results in the case where X and Z are independent: SD1=$\sqrt{\frac{1}{1000}\sum_j \hat{V}(\hat{\beta})}$, SD2=$\sqrt{V(\hat{\beta})}$, NA=fraction not converged.

			Frailty Model				***Stratified Model***			
n	β	Dist	Bias	SD1	SD2	NA	Bias	SD1	SD2	NA
300	1.2	Gamma	0.783	0.144	0.150	0.000	0.007	0.260	0.272	0.000
75	1.2	Gamma	0.772	0.285	0.280	0.004	0.015	0.549	0.554	0.000
300	1.2	Log-Normal	0.628	0.128	0.134	0.000	0.010	0.251	0.253	0.000
75	1.2	Log-Normal	0.631	0.258	0.264	0.000	0.031	0.528	0.539	0.002

Table 2: Simulation results in the case where X and Z are dependent, SD1=$\sqrt{\frac{1}{1000}\sum_j \hat{V}(\hat{\beta})}$, SD2=$\sqrt{V(\hat{\beta})}$, NA=fraction not converged.

independent of Z), in the second case it is drawn from a Bernoulli distribution with a probability which is determined by Z_i to yield a correlation between X and Z which is approximately 0.7.

We simulate for two different values of n (e.g. number of twin pairs), $n = 300, 75$ and all scenarios are replicated 1000 times. In each case, the value of $\beta_0 = \log(1.2)$. In each run, the results from fitting the correct frailty model (either Γ or log-norma), and a stratified Cox model are saved. The waiting times, T_{ij}, are censored by the variable U_{ij} drawn from an exponential with a rate of 0.005 giving rise to approximately $40 - 45\%$ of the twin pairs having the shortest waiting time censored (these pairs do not contribute to the partial likelihood, cf. (3)) The simulations are run in **R**.

References

Andersen, PK, Borgan, Ø, Gill, RD, Keiding, N (1993). *Statistical Models Based on Counting Processes.* New York: Springer-Verlag.

Holt, JD, Prentice, RL (1974). Survival analysis in twin studies and matched pair experiments. *Biometrika* **61**, 17-30.

Madsen, M, Andersen PK, Gerster M, Andersen A-MN, Christensen K, Osler, M (2011). Education and incidence of breast cancer: Does the association replicate within twin pairs? A register based study on Danish female twins. *Br. J. Cancer* **104**, 520-523.