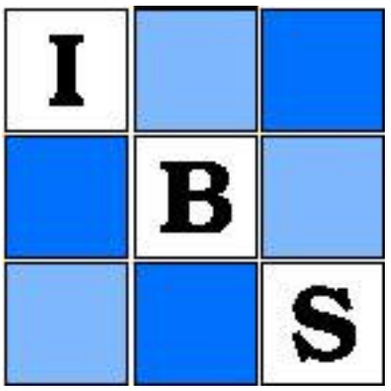

BIOMETRICS 51, 524–532
June 1995

Applying Cox Regression to Competing Risks

Mary Lunn[1] and Don McNeil[2]

[1]St. Hughs College, Oxford, OX26LE, United Kingdom
[2]Macquarie University & NHMRC Clinical Trials Centre, the University of Sydney, NSW 2006, Australia

SUMMARY

Two methods are given for the joint estimation of parameters in models for competing risks in survival analysis. In both cases Cox's proportional hazards regression model is fitted using a data duplication method. In principle either method can be used for any number of different failure types, assuming independent risks. Advantages of the augmented data approach are that it limits over-parametrisation and it runs immediately on existing software. The methods are used to reanalyse data from two well-known published studies, providing new insights.

1. Introduction

In trials involving survival data it is often the case that there are competing risks involved, so that one needs to assign the type of failure, say type I or II, in addition to failed/censored status. One objective of practical interest is to determine the effect of removing a cause of failure on the survival distribution, a problem investigated as long ago as 1760 by Daniel Bernoulli (see, for example, David and Moeschberger (1978)). More recent contributions to the subject have come from Kalbfleisch and Prentice (1980) and Cox and Oakes (1984).

Two approaches have been followed for analysing cause-specific survival data. The first, described by Kalbfleisch and Prentice (1980) and used, for example, by Kay (1986), simply involves fitting models separately for each type of failure in turn, treating other failure types as censored data. A drawback in this method as it stands is that it does not treat the different types of failures jointly, complicating the comparison of parameter estimates corresponding to different failure types. An alternative approach, which has been used by various authors including Larson and Dinse (1985) and Kuk (1992), involves fitting more complex models incorporating the different failure types. A difficulty with this alternative is that standard software is not available: Kuk (1992) needed to use Monte Carlo simulation to accommodate two failure types in a survival analysis model.

In this paper we demonstrate that it is possible to analyse competing risks in survival analysis using readily available standard programs for fitting Cox's (1972) proportional hazards regression model with censored data. Assuming initially that there are just two failure types in addition to censoring, we will show that by augmenting the data using a duplication method Cox regression can be adapted in either of two ways to take account of the failure types. Two vectors of regression coefficients $\mathbf{b}_{\mathrm{I}}$ and $\mathbf{b}_{\mathrm{II}}$ may be defined depending on the type of failure. One procedure runs Cox regression stratified by type of failure. The other procedure uses unstratified Cox regression, assuming that the hazard functions associated with the two types of failure have a constant ratio. Generally this is not likely to be the case, but often a good fit can be found by separating the model into different time zones. Advantages of the augmented data approach are that it does not overparametrise the model and it can be run immediately using existing software, thus exploiting tests within standard statistical packages for concluding whether or not the regression coefficients are the same for each failure type.

2. The Methods

Consider first the data layout. For the moment assume two failure types. The data for each individual subject should be entered twice, the second entry shown as the other type of failure but

Key words: Censoring; Cox regression; Data duplication; Failure type; Independent competing risks.

always censored. Suppose that types I and II are given by $\delta = 0$ or 1, respectively. For example suppose that subject **i** fails at time t_i and has failure type δ_i. We make the following two entries.

Subject	Response	Status	Failure type	Covariates
i	t_i	1	δ_i	$\mathbf{x}_i$, $\delta_i\mathbf{x}_i$
i (rep)	t_i	0	$1 - \delta_i$	$\mathbf{x}_i$, $(1 - \delta_i)\mathbf{x}_i$

If the subject is censored we again make two entries, as above, but with each showing a censored observation, one for each failure type. We then regress on the duplicated covariates. The first method uses δ as a covariate, the second uses it as a stratifying variable.

The basis for the data duplication is as follows. The hazard functions for the two types of risk are assumed to be additive. Thus the hazard of failure is the sum of two component risk processes, and the time to failure of either type is the minimum of the two failure times associated with these risk processes. When observation ceases due to failure or censoring, two survival times of the same duration have been observed, one for each process, at least one being censored.

The covariates $\mathbf{x}_i$ are augmented to allow for possible interactions with type of failure.

2.1 *Method A*

Cox regression is run on the augmented data set, failure type δ being included with the covariates x and δx. Assuming no ties the contribution to the partial likelihood if observation **i** results in a failure is

$$e^{b_0\delta_i + b'x_i + \theta'\delta_i x_i} \Big/ \sum_{R_i} e^{b_0\delta + b'x + \theta'\delta x}, \tag{1}$$

where the summation is over all survival times (including each appropriate second entry) which have neither failed nor been censored at time t_i. Note that the covariates for each entry are either of the form (**x**, 0) or of the form (**x**, **x**), so that the contribution to the denominator from the two duplicated entries with original covariate **x** is

$$e^{b'x} + e^{b_0 + b'x + \theta'x}.$$

Essentially this means that we are assuming that the hazard functions for the two failure types for an individual with covariates **x**, are proportional to

$$e^{b'x},\ e^{b_0 + b'x + \theta'x},$$

with the same baseline hazard function $\lambda_{01}(t)$. The hazard function for risk type I of a subject with covariates **x** is thus $\lambda_{01}(t)\exp(b'x)$. The hazard function for risk type II for the same subject is $\lambda_{02}(t)\exp(b'x + \theta'x) = \lambda_{01}(t)\exp(b_0 + b'x + \theta'x)$. The baseline hazard functions of the two types (corresponding to $\mathbf{x} = 0$) differ by a constant ratio $\exp(b_0)$. Care should be taken in interpreting the coefficient b_0 as it is not invariant under change of location of the covariate vector **x**. Note that neither is the function $\lambda_{01}(t)$. The expression (1) above is the probability that subject **i** has a failure type δ, from all those subjects available for failure of either type.

For ease of analysis transform the regression coefficients to $b_{\mathrm{I}} = b$, $b_{\mathrm{II}} = b + \theta$, these being the vectors of regression coefficients appropriate to each type of failure. In the absence of ties the full partial log-likelihood becomes

$$L = \sum_{j,\mathrm{I}} b_{\mathrm{I}}'x_j + \sum_{j,\mathrm{II}} (b_0 + b_{\mathrm{II}}'x_j) - \sum_j \ln\left[\sum_{R_j} (e^{b_{\mathrm{I}}'x} + e^{b_0 + b_{\mathrm{II}}'x})\right], \tag{2}$$

where the first sum is over the type I failures, the second is over the type II failures, and the last sum is over all failures. The sum inside the brackets is taken over all covariates of observations in the risk set R_j. The equations to give maximum partial likelihood estimates of b_{I}, b_{II}, and b_0 are of the form

$$\frac{\partial L}{\partial b_{\mathrm{I}}} = 0 = \sum_{j,\mathrm{I}} x_j - \sum_j \left[\sum_{R_j} xe^{b_{\mathrm{I}}'x} \Big/ \sum_{R_j} (e^{b_{\mathrm{I}}'x} + e^{b_0 + b_{\mathrm{II}}'x})\right]$$

$$\frac{\partial L}{\partial b_{\mathrm{II}}} = 0 = \sum_{j,\mathrm{II}} x_j - \sum_j \left[\sum_{R_j} xe^{b_0 + b_{\mathrm{II}}'x} \Big/ \sum_{R_j} (e^{b_{\mathrm{I}}'x} + e^{b_0 + b_{\mathrm{II}}'x})\right] \tag{3}$$

$$\frac{\partial L}{\partial b_0} = 0 = \sum_{\text{II}} 1 - \sum_j \left[\sum_{R_j} e^{b_0 + b'_{\text{II}} x} \Big/ \sum_{R_j} (e^{b'_{\text{I}} x} + e^{b_0 + b'_{\text{II}} x}) \right].$$

Running standard Cox regression on the augmented data set gives the appropriate estimates of the regression coefficients for each type and also the appropriate P values, provided the model fit is good. With a moderate fit the robust estimates of Lin and Wei can be derived, again from the augmented data set.

The regression coefficient θ represents the difference between the two vectors of regression coefficients for the failure types. Standard P values will indicate the statistical significance of the components of θ.

In the extreme case where there are no type II failures then the third equation in (3) above gives $b_0 = -\infty$ and the regression becomes the standard regression with one failure type only.

The partial likelihood which results from Method A is precisely the partial likelihood suggested by Kalbfleisch and Prentice (1980) for competing risks with baseline hazard functions differing by a constant ratio.

2.2 *Method B*

Run a Cox regression on the covariates $\mathbf{x}$, $\delta\mathbf{x}$ stratifying by failure type, $\delta = 0$ or 1. In this case the partial likelihood is

$$\prod_{t_i, \delta_i = 0} \left(\frac{e^{b' x_i}}{\sum_{R_i} e^{b' x}} \right) \prod_{t_i, \delta_i = 1} \left(\frac{e^{b' x_i + \theta x_i}}{\sum_{R_i} e^{b' x + \theta' x}} \right), \tag{4}$$

treating the survival times of the two types of failure separately. In each case the risk set R_i consists of those subjects with the appropriate stratum identifier, $\delta = 0$ for the first product and $\delta = 1$ for the second. This is the partial likelihood for two failure types with regression coefficients b, $b + \theta$ and unknown baseline hazard functions h_{00}, h_{01} whose relationship h_{01}/h_{00} is not known. The test $\theta_k = 0$ routinely applied in statistical packages becomes a test of whether or not regression on the kth covariate is the same for each failure type.

The difference in the two methods lies in the assumption in Method A that there is a constant ratio h_{01}/h_{00} of the baseline hazard functions. If this is the case then method A can be expected to be more powerful than Method B. If not then stratifying the data by time zone and using Method A in each zone may produce a good fit. Whichever method of estimating the parameters is used, standard methods can then be employed to find estimated survival curves and absolute risks (Benichou and Gail, 1990) also residuals (Barlow and Prentice, 1988). In Section 3 we give robust estimates for standard errors of regression coefficients following Lin and Wei (1989) and Wei, Lin, and Weissfeld (1989).

It is clear that the methods can be developed to account for $J > 2$ failure types. The data are duplicated J times, one row for each failure type. In Method A we will require $J - 1$ indicator variables so that $\delta_k = 1$ if failure type $k + 1$ occurs. In Method B we stratify by failure type.

3. Stanford Heart Transplant Data

As an illustration of the methods, consider survival times of 65 patients who received heart transplants in the Stanford Heart Project up to 1 April 1974, reported by Crowley and Hu (1977). The failures are classified by death from rejection or other cause. These data have been considered by many authors, including Larson and Dinse (1985) and Kuk (1992) who modelled the causes of death in competing risk terms.

Using the data augmentation method, Kaplan–Meier curves may be constructed for each cause in the absence of the other, and these are shown together with the standard Kaplan–Meier curve in Figure 1. The cause-specific curves do not have a true survivorship interpretation, but represent $\exp(-H_t)$, where H_t is the cumulative cause-specific hazard function to time t, for each failure type (see, for example, Kalbfleisch and Prentice (1980), page 168). Since the cause-specific curves cross, it is not expected that the proportional hazards model will provide a satisfactory fit.

Method A involves fitting the proportional hazards model, the covariates being type of failure (δ), age at transplant minus the median age of 48 years (x_1), and tissue mismatch score (x_2), together with their interactions δx_1 and δx_2; the result is shown in Table 1. P values for the test of the proportional hazards assumption suggested by Harrell and Lee (1986), denoted by p(PH), are small for δ and the two interactions, confirming the poor fit of the model. Table 2 gives the results from fitting the reduced model in which the parameters corresponding to the deaths from other causes are omitted.

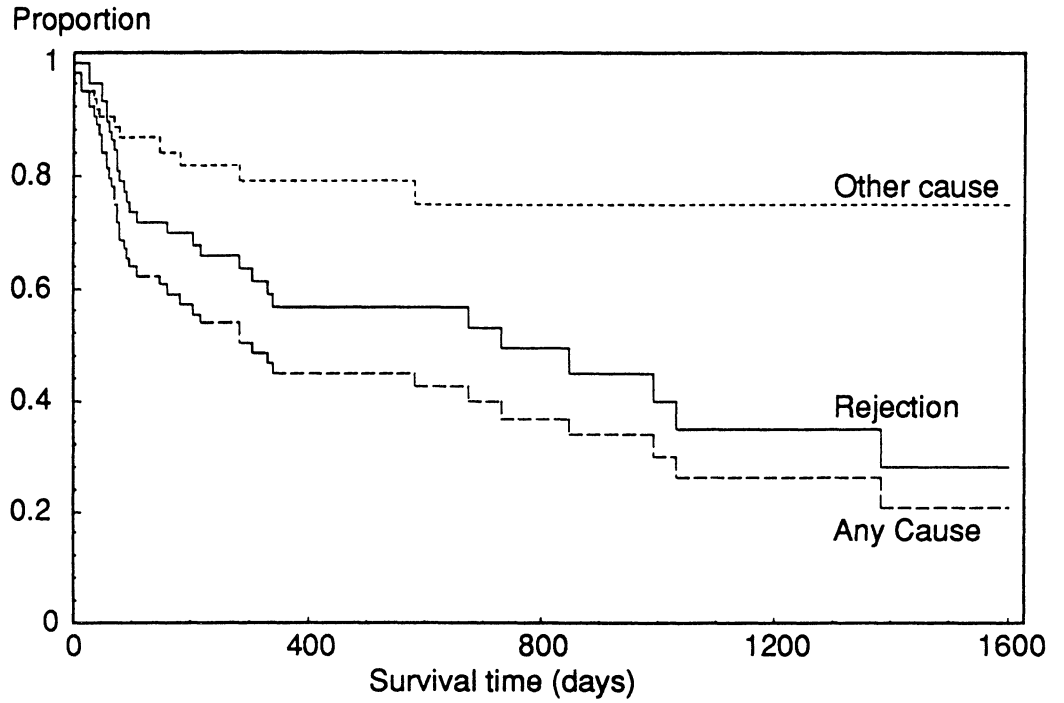

Figure 1. Cause-specific and overall Kaplan–Meier curves of heart transplant survival

Table 1
Result of fitting cause-specific proportional hazards model to heart transplant data

Covariate	Coeff	SE[a]	*P* value	*p*(PH)
Reject	−.196	.788		.003
Age at transplant	−.010	.034	.768	.127
Mismatch score	−.237	.570	.678	.217
Reject.age[b]	.107	.046	.020	.175
Reject.MMscore[b]	.978	.640	.126	.007

Partial likelihood (log times −2): 323.31.
[a] SE, standard error.
[b] Interaction terms, for example age by reject failure type.

Table 2
Reduced cause-specific proportional hazards model: Heart transplant data

Covariate	Coeff	SE	*P* value	*p*(PH)	Robust SE
Reject	.012	.529		.004	.503
Reject.age	.098	.032	.002	.176	.023
Reject.MMscore	.745	.303	.014	.008	.296

Partial likelihood (log times −2): 323.57.

Following Lin and Wei (1989) robust standard errors may be computed, and these are also given in the rightmost column of Table 2.

Method B involves fitting a stratified proportional hazards model, the stratification variable being failure type (δ) and the covariates being x_1 and x_2, together with the interactions δx_1 and δx_2, as before. The result from fitting this model is shown in Table 3.

For this model the covariates could be recoded as the pairs δx_1, δx_2 and $(1 - \delta)x_1$, $(1 - \delta)x_2$, instead of the equivalent set of main effects and interactions, and the resulting fit is shown in Table 4. With the parameters recoded in this way, the result is identical to that obtained by fitting proportional hazards models separately for the two failure types (treating failures from the other type as censored data). The reason for the equivalence is that with this reparameterisation the likelihood (4) may be rewritten as the product of two components for the two failure types, each

Table 3
Cause-specific stratified proportional hazards model: Heart transplant data

Covariate	Coeff	SE	*P* value	*p*(PH)
Age at transplant	−.012	.032	.702	.395
Mismatch score	−.322	.532	.525	.369
Reject.age	.121	.046	.009	.324
Reject.MMscore	1.227	.623	.049	.155

Partial likelihood (log times −2): 270.54.

Table 4
Reparameterised cause-specific stratified proportional hazards model

Covariate	Coeff	SE	*P* value	*p*(PH)	Robust SE
Reject.age	.109	.034	.001	.323	.028
Reject.MMscore	.905	.323	.005	.155	.346
Other.age	−.012	.032	.702	.495	.033
Other.MMscore	−.322	.532	.525	.292	.739

Partial likelihood (log times −2): 270.54.

involving independent pairs of parameters, so it is maximised when each component is maximised. Note that the age and mismatch effects are only significant for the rejection failures.

The *P* values for the interaction terms in the joint model given in Table 3 give the results of the tests that the two sets of parameters are the same. Using the estimated covariance matrix obtained from robust estimation as in Wei et al. (1989) robust standard errors are obtained, as given in Table 4. If robust estimates are used for the covariance matrix then the *P* values testing whether the two sets of parameters for age and mismatch score are the same in type I and type II failures become .004 for age and .198 for mismatch score. The latter result follows because the robust estimates suggest a correlation of −.468 between the estimated regression coefficients for mismatch score. However the separate *P* values on the estimated regression coefficients for mismatch score show no significant change using the robust estimates.

A serious drawback of the stratified model is that it does not fit parameters to the stratification variable, and thus does not allow the effect of this variable to be estimated and tested statistically. An alternative method which does allow modelling of covariates failing the proportional hazards assumption involves incorporating time dependence into the covariates.

Larson and Dinse (1985) suggested fitting a model to these data in which the relative risks corresponding to the covariates are step functions constant on the three intervals 0–45 days, 45–90 days, and 90+ days. They found that the covariates x_1 and x_2 were only risk factors for the patients dying from transplant rejection. However with only 12 subjects dying from causes other than rejection, there is little information available to study time dependence of the various effects. With this cautionary note, it is instructive to extend the model based on Method B to allow for time dependence.

The proportional hazards model may be fitted separately on intervals by selecting the subjects to survive to the beginning of the interval and censoring all survival times at the end of the interval. However, it is possible to combine these results using a method outlined by Harrell and Lee (1986), and fitting this model, with just two time zones divided at 60 days, gives the output in Table 5. The effects of age and mismatch score for the non-rejection deaths have been omitted, so that the model extends the simple proportional hazards model given in Table 2. Note that the difference in deviance between the two comparable models is 8.87 with only 3 degrees of freedom, so there is a significant improvement in fit.

Examining this result, two parameter estimates are highly significant and no others come close to being statistically significant, so the final model may be interpreted as follows. The risks of death from rejection and from other causes are not significantly different from each other in the first 60 days. After 60 days the logarithm of the ratio of the risk of death from rejection or from other causes increases linearly with age and with mismatch score. During this period a typical patient aged 48 with a mismatch score of 0 has the same risk of dying from rejection as from other causes, but a unit increase in mismatch score increases this relative risk by a factor of exp(.961), or 2.61, and an age increase of ten years increases it by a factor of exp(1.32), or 3.74.

Table 5
Cause-specific model for transplant data, with hazards proportional on two intervals

	0–59 days			60+ days		
Covariate	Coeff	SE	*P* value	Coeff	SE	*P* value
Reject	−.903	1.116		.271	.660	
Reject.age	.027	.058	.642	.132	.040	.001
Reject.MMScore	.583	.664	.379	.961	.364	.008

Partial likelihood (log times −2): 314.70.

4. Prostate Cancer Data

As a second illustration, consider the survival times of 506 patients with prostate cancer who entered a clinical trial during 1967–1969 and were randomly allocated to different levels of treatment with the drug diethylstilbestrol (DES). These data were considered by Byar and Green (1980) and are published in full in Andrews and Herzberg (1985). Considering eight specified risk factors including drug treatment (defined as at least 1 mg of DES daily) and omitting 23 subjects with incomplete data, Kay (1986) fitted separate cause-specific Cox proportional hazards models, classifying the causes of death as 1) cancer, 2) cardiovascular (CVD), or 3) other.

While Kay's approach enables the various risk factors to be evaluated for each cause, comparison of the risk factors across causes is complicated by the fact that the model survival curves are cause-specific. Figure 2 shows the overall and cause-specific Kaplan–Meier curves, the interpretation being similar to the curves shown in Figure 1.

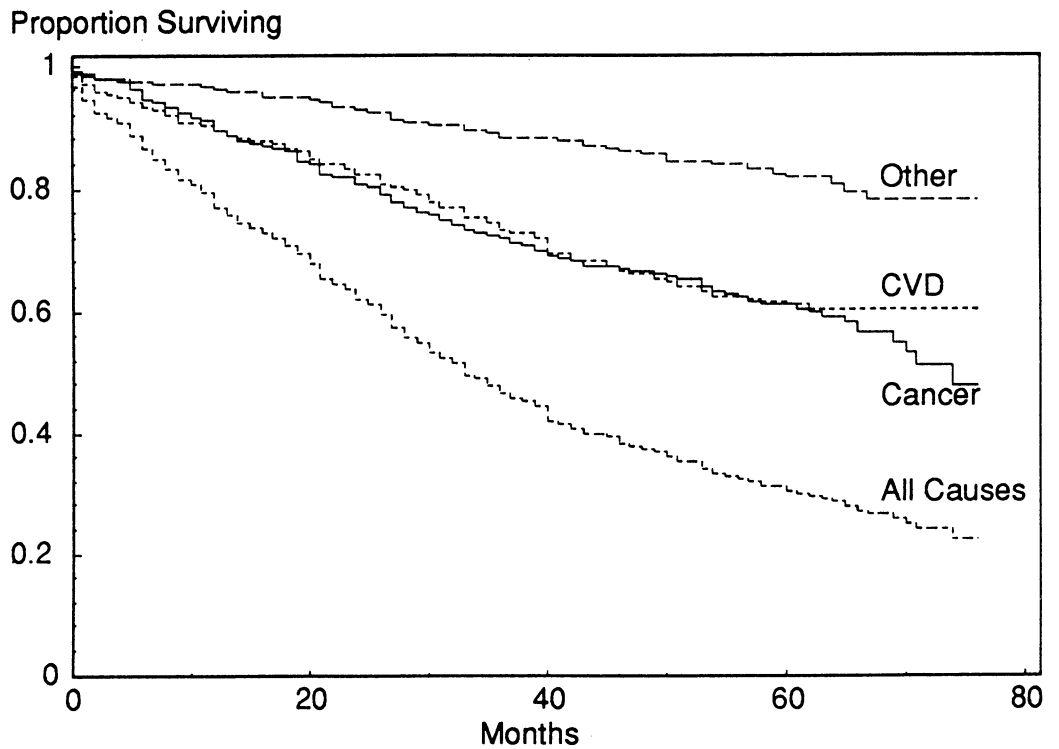

Figure 2. Cause-specific survival curves for 483 patients with prostate cancer

The model fitted by Kay is identical to that obtained using Method B, provided the interaction terms between each risk factor and cause of death are included. Since the survival curves cross (the risk of CVD death being relatively high at first), it is not reasonable to use Method A for these data. However if a further 43 subjects with survival times less than 5 months are excluded, the full proportional hazards assumption is tenable. Figure 3 shows estimated cause-specific curves for the 440 patients who survived for 5 months or more. The estimates of the cause-specific survival proportions, p_i, are re-expressed as $-\ln[-\ln(p_i)]$ to facilitate visual assessment of the proportional

hazards assumption: if the assumption is valid, the re-expressed curves should be (approximately) parallel.

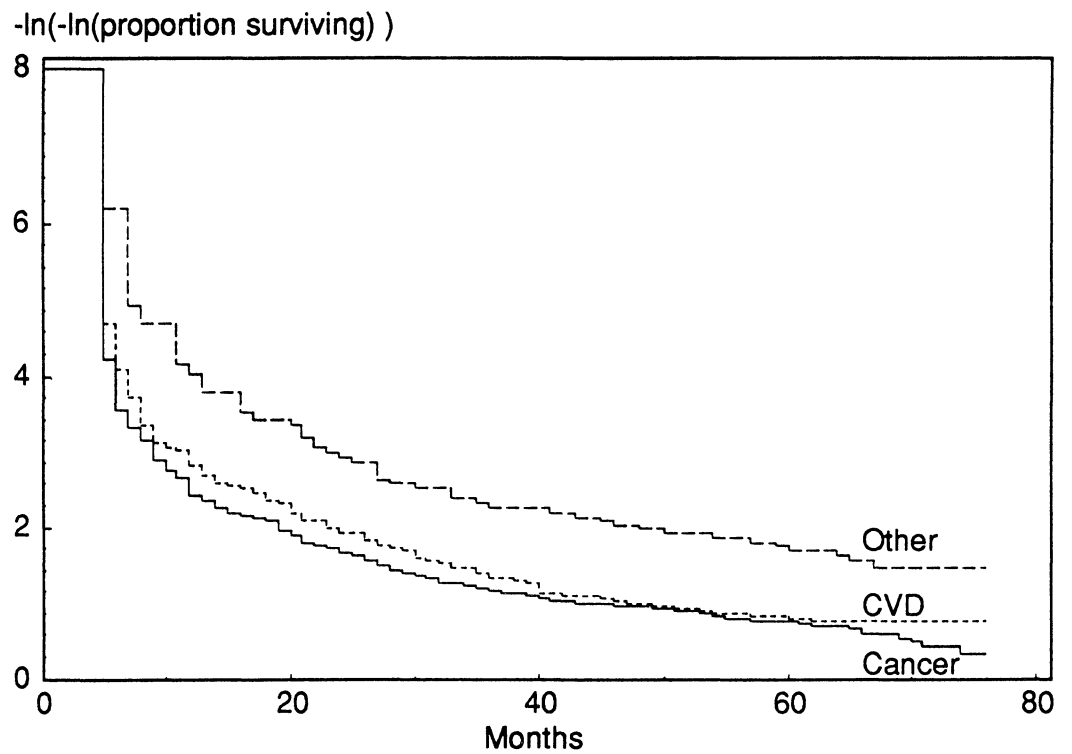

Figure 3. Re-expressed cause-specific Kaplan–Meier curves for 440 patients with prostate cancer surviving 5 months or more

Since the curves in Figure 3 are nearly parallel, it is reasonable to use Method A to model the reduced set of data, and the results are given in Tables 6 and 7. Table 6 gives the fit from a simple model for cause-specific failure with no covariates. In this model the two coefficients reflect the relative risks of death from (i) CVD and (ii) other causes, respectively, relative to cancer. Thus the model shows that the risks of death from CVD and cancer are not significantly different, but the risk of death from other causes is approximately one third of the risk of death from cancer. These results match Figure 2.

Table 6
Simple cause-specific PH model with proportional baseline hazard functions (Method A): prostate cancer patients surviving for 5 months or more

Cause	Coeff	SE	*P* value	*p*(PH)
CVD-cancer[c]	−.174	.126	.168	.323
Oth-cancer[c]	−1.077	.169	.000	.080

Partial likelihood (log times −2): 3964.65.
[c] Relative hazard uses cancer hazard as baseline.

Table 7 gives the parameter estimates after fitting a more complex model which includes treatment and other risk factors for each of the three causes of death. The risk factors, labelled by AG, WT, PF, HX, HG, SZ, and SG, correspond, respectively, to age group, weight index, performance rating, history of cardiovascular disease, serum haemoglobin, size of primary lesion, and Gleason stage/grade category. The improvement in the fit of the complex model over the simple model is quite substantial: the 24 additional parameters decrease the deviance by 195.61.

There are two advantages in using Method A in this case. First, the non-stratified proportional hazards model has a single underlying survival curve, so that all of the coefficients in the model are

Table 7
Complex cause-specific PH model with proportional baseline hazard functions (Method A): prostate cancer patients surviving for 5 months or more

	Cancer			CVD			Other		
Covariate	Coeff	SE	p(PH)	Coeff	SE	p(PH)	Coeff	SE	p(PH)
Cause-cancer				−.225	.325	.146	−.554	.391	.162
Treatment	−.564	.177	.367	.330	.190	.139	−.505	.302	.207
AG	−.061	.151	.358	.369	.149	.280	.821	.224	.367
WT	.209	.143	.065	.096	.164	.357	.509	.252	.429
PF	.270	.274	.306	.481	.307	.484	−.074	.608	.497
HX	−.063	.186	.285	1.104	.201	.384	.150	.307	.006
HG	.465	.185	.162	−.061	.236	.070	−.243	.411	.232
SZ	1.131	.214	.040	−.305	.438	.291	.731	.489	.312
SG	1.335	.206	.353	−.047	.203	.236	−.611	.336	.474

Partial likelihood (log times −2): 3779.04.

directly comparable in terms of relative risks. Thus, for example, Table 7 reveals that (other things being equal) the treatment might reduce the risk of a cancer death at any time by a factor exp(−.564), or .57, but might increase the risk of a cardiovascular death by a factor exp(.330), or 1.39. The stratified analysis does not allow such a simple interpretation. The second advantage of the method is that it is possible to proceed to fit more easily interpretable reduced models in which parameters are constrained. For example the two constant coefficients given in Table 7 could be eliminated from the model.

Acknowledgements

We thank John Simes and the two referees for perceptive comments which clarified the discussion. The analysis was undertaken and the graphs produced using the statistical package SPIDA.

Résumé

Deux méthodes sont présentées pour l'estimation jointe des paramètres d'un modèle de survie à risques compétitifs. Dans les deux méthodes, le modèle de régression à risques proportionnels de Cox est ajusté en utilisant une méthode de duplication des données. Ces méthodes penvent en principe être appliquées quelque soit le nombre de causes d'échec, en faisant l'hypothèse d'indépendance des différents risques. Cette approche d'augmentation des données a pour avantage de limiter la surparamétrisation et de pouvoir être utilisée sur tous les logiciels existants. Ces méthodes ont été appliquées aux données de deux études publiées connues, auxquelles elles apportent quelques aperçus nouveaux.

References

Andrews, D. F. and Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker*. New York: Springer-Verlag.

Barlow, W. E. and Prentice, R. L. (1988). Residuals for relative risk regression. *Biometrika* **75,** 65–74.

Benichou, J. and Gail, M. H. (1990). Estimates of absolute cause-specific risk in cohort studies. *Biometrics* **46,** 813–826.

Byar, D. P. and Green, S. B. (1980). Prognostic variables for survival in a randomized comparison of treatments for prostatic cancer. *Bulletin du Cancer (Paris)* **67,** 477–490.

Crowley, J. and Hu, M. (1977). Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* **72,** 27–36.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34,** 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

David, H. A. and Moeschberger, M. L. (1978). *The Theory of Competing Risks*. London: Griffin.

Harrell, F. and Lee, K. (1986). Verifying assumptions of the proportional hazards model. *Proceedings of the Eleventh Annual SAS User's Group International*, pp. 823–828.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons.

Kay, R. (1986). Treatment effects in competing-risks analysis of prostate cancer data, *Biometrics* **42,** 203–211.

Kuk, A. Y. C. (1992). A semiparametric mixture model for the analysis of competing risks data. *Australian Journal of Statistics* **34,** 169–180.

Larson, M. G. and Dinse, G. E. (1985). A mixture model for the regression analysis of competing risks data. *Applied Statistics* **34,** 201–211.

Lin, D. Y. and Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association* **84,** 1074–1079.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* **84,** 1065–1073.

Received January 1993; *revised July* 1993 *and May* 1994; *accepted May* 1994.