BIOMETRICS 47, 461-466
June 1991

Regression with Frailty in Survival Analysis

C. A. McGilchrist and C. W. Aisbett

University of New South Wales,
P.O. Box 1, Kensington N.S.W. 2033, Australia

SUMMARY

In studies of survival, the hazard function for each individual may depend on observed risk variables but usually not all such variables are known or measurable. This unknown factor of the hazard function is usually termed the individual heterogeneity or frailty. When *survival* is time to the occurrence of a particular type of event and more than one such time may be obtained for each individual, frailty is a common factor among such recurrence times. A model including frailty is fitted to such repeated measures of recurrence times.

1. Introduction

In an experiment, each of M individuals is observed and times between recurrences of a particular type of event are recorded. The problem that motivates this study is the recurrence of infection in kidney patients who are using a portable dialysis machine. The infection occurs at the point of insertion of the catheter and, when it occurs, the catheter must be removed, the infection cleared up, and then the catheter reinserted. Recurrence times are times from insertion until the next infection. Sometimes the catheter must be removed for other reasons so that there may be right censoring of the data. As well, the final recurrence time may be censored. It is assumed here that each patient is followed for a predetermined number of recurrence times, some of which may be censored.

The process is modelled through the hazard function of infection, which may depend on several risk variables collected into a vector $\mathbf{x}$. The hazard function is denoted by $h(t, \mathbf{x})$ where t is the time from the beginning of the current interval. A proportional hazards model for this function gives

$$h(t,\mathbf{x}) = \lambda(t)\,g(\mathbf{x}),$$

where $\lambda(t)$ describes the variation of hazard over time and $g(\mathbf{x})$ is the multiplicative effect of the combined risk variables on this function. The model is extended to include a multiplicative individual heterogeneity or frailty term Z in

$$h(t,\mathbf{x}) = Z\lambda(t)\,g(\mathbf{x}).$$

The recurrence times for each individual have a common frailty that may be regarded as a random selection from some suitably defined population distribution of frailties. Because a 10-week interval is allowed between an infection and reinsertion of the catheter, the recurrence intervals are taken to be independent apart from their common frailty component.

The variables recorded for each individual are denoted here by

i = Patient number, $\quad i = 1, 2, \ldots, M$;

T_{ij} = jth smallest recurrence time for patient i, $\quad j = 1, 2, \ldots, m_i$;

$\mathbf{x}_{ij}$ = Vector of risk variables applying to patient i at his jth ordered recurrence time.

Key words: Frailty; Hazard functions; Survival analysis.

There are N recurrence times and the ordered recurrence times, which are assumed here to be distinct, are denoted by t_n. Let

$$D_{in} = \begin{cases} 1 & \text{if an } \textit{event} \text{ occurs to individual } i \text{ at } t_n, \\ 0 & \text{otherwise.} \end{cases}$$

The basic problem is to estimate parameters of the regression function $g(\mathbf{x})$. Although the Cox (1972) model $g(\mathbf{x}) = \exp(\mathbf{x}'\boldsymbol{\beta})$ is later applied, a general function g is retained in the exposition but usually written as $g(\mathbf{x}, \boldsymbol{\beta})$ to indicate its dependence on a vector parameter $\boldsymbol{\beta}$. There is a vast literature on the estimation of such models without the inclusion of a frailty term. The treatment of frailties has been dealt with by Clayton and Cuzick (1985), Oakes (1982), Crowder (1985), Holt and Prentice (1974), Prentice, Williams, and Peterson (1981), and Hougaard (1984, 1986a, 1986b). The current approach is similar to the Clayton and Cuzick approach in that it replaces the random frailty terms by shrinkage estimates. Using posterior modes instead of posterior means gives a computationally simpler procedure that reverts as much as possible to the original Cox solution.

2. Estimation

When event times are ordered and individual i has m_i such times, the first occurring event is the smallest order statistic for that individual. The corresponding hazard rate is that which applies to the smallest order statistic. Using Z_i for the frailty of individual i, a general expression for the hazard that he has an event at t_n, given knowledge of events up to that time, is

$$Z_i \lambda(t_n) a_{in}(\boldsymbol{\beta}), \quad \text{where} \quad a_{in}(\boldsymbol{\beta}) = \sum_{j=r_{in}}^{m_i} g(\mathbf{x}_{ij}).$$

Summations with respect to j when $m_i < r_{in}$ are taken to be zero and r_{in} is 1 plus the number of events or censoring times occurring to individual i before time t_n. This type of hazard function for order statistics has been used by Gail, Santner, and Brown (1981) in the modelling of multiple times to tumour.

In the construction of a partial likelihood (Cox, 1975) for the data, an individual remains in the risk set for times at or before his longest recurrence interval. Thus the logarithm of the partial likelihood is

$$\ell = \sum_{n=1}^{N} \sum_{i=1}^{M} D_{in}[\ln Z_i + \ln b_{in}(\mathbf{z}, \boldsymbol{\beta})] = \sum_{n=1}^{N} \sum_{i=1}^{M} D_{in} \ln(Z_i b_{in}),$$

where $b_{kn}(\mathbf{z}, \beta) = a_{kn} / \sum_i Z_i a_{in}$. In what follows it is easier to estimate $U_i = \ln Z_i$ rather than the Z_i, which are restricted to be positive. The derivatives of ℓ with respect to $\boldsymbol{\beta}$ and U_k may be written

$$\partial \ell / \partial \boldsymbol{\beta} = \sum_{n=1}^{N} \sum_{i=1}^{M} D_{in} \partial \ln b_{in} / \partial \boldsymbol{\beta},$$

$$\partial \ell / \partial U_k = \sum_{n=1}^{N} \left\{ D_{kn} - \sum_{i=1}^{M} D_{in} Z_k b_{kn} \right\} = D_{k\cdot} - \sum_{n=1}^{N} D_{\cdot n} Z_k b_{kn},$$

where $D_{k\cdot}$ is the total number of events for individual k; $D_{\cdot n}$ is 1 if an event occurs at time t_n, otherwise it is zero. Here we have used $\partial b_{in} / \partial Z_k = -b_{in} b_{kn}$. The second-order derivatives may be readily obtained as

$$\partial^2 \ell / \partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}' = \sum_{n=1}^{N} \sum_{i=1}^{M} D_{in} \partial^2 \ln b_{in} / \partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}',$$

$$\partial^2 \ell / \partial \boldsymbol{\beta}\, \partial U_k = - \sum_{n=1}^{N} D_{\cdot n} b_{kn} Z_k \partial \ln b_{kn} / \partial \boldsymbol{\beta},$$

$$\partial^2 \ell / \partial U_k\, \partial U_{k'} = - \sum_{n=1}^{N} D_{\cdot n} [Z_k b_{kn} \delta_{kk'} - Z_k b_{kn} Z_{k'} b_{k'n}],$$

where $\delta_{kk'}$ is the usual Kronecker delta.

Because frailties are merely relative to one another, the location parameter of the distribution of log frailties may be arbitrarily fixed to zero. Maximum likelihood estimation of the U_k is therefore subject to the restriction $\sum U_k = 0$.

A Newton–Raphson scheme for iterative estimation of $\boldsymbol{\beta}$ and the vector of frailties $\mathbf{u}$ begins with initial estimates $\boldsymbol{\beta}_0, \mathbf{u}_0$, where $\mathbf{u}_0$ must satisfy $\mathbf{1}'\mathbf{u}_0 = 0$, i.e., all components of $\mathbf{u}_0$ must sum to zero. An approximation to $\hat{\boldsymbol{\beta}}, \hat{\mathbf{u}}$ is

$$\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}_0 \\ \mathbf{u}_0 \end{bmatrix} + \mathbf{V}^{-} \begin{bmatrix} \partial \ell / \partial \boldsymbol{\beta} \\ \partial \ell / \partial \mathbf{u} \end{bmatrix}_{\boldsymbol{\beta} = \boldsymbol{\beta}_0,\, \mathbf{u} = \mathbf{u}_0},$$

where $\mathbf{V}$ is the estimated information matrix

$$- \begin{bmatrix} \partial^2 \ell / \partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}' & \partial^2 \ell / \partial \boldsymbol{\beta}\, \partial \mathbf{u}' \\ \partial^2 \ell / \partial \mathbf{u}\, \partial \boldsymbol{\beta}' & \partial^2 \ell / \partial \mathbf{u}\, \partial \mathbf{u}' \end{bmatrix}.$$

The generalised inverse $\mathbf{V}^{-}$ is chosen to be orthogonal to $[\mathbf{0}', \mathbf{1}']$ and hence the solution $\hat{\mathbf{u}}$ continues to satisfy the restriction $\mathbf{1}'\hat{\mathbf{u}} = 0$. Such a technique is equivalent to the use of Lagrange multipliers as described in Aitchison and Silvey (1959).

This Newton–Raphson procedure may converge if there is sufficient variation of the measured risk variables within each patient. However, if the major variation of risk variables is from patient to patient rather than within each patient, there is a fundamental lack of identifiability in the problem. The variation of hazard rate may be attributed to dependence on the risk variables or dependence on the frailty terms and there is a consequent failure to converge. A partial resolution of the problem is achieved, as is common in experimental design problems, by considering the frailty as a random component. This is done in the next section.

3. Distribution of Frailties

One method of including the random selection of the frailty parameters in the estimation process is to combine a prior frailty distribution with maximum likelihood estimates obtained in Section 2. This approach leads to ridge regression of Hoerl and Kennard (1970) or preshrunk estimators of Copas (1983). It might also be called a *penalised partial likelihood* but perhaps the oldest similar technique is the best linear unbiased predictor (BLUP) method of Henderson (1963, 1973, 1975). Suppose the log frailties are distributed normally with zero mean and unknown variance. Because ln(frailties) must sum to zero and the prior distribution should be symmetric in $\mathbf{u}$, the variance matrix for $\mathbf{u}$ is taken to be $\sigma^2(\mathbf{I} - \mathbf{M}^{-1}\mathbf{1}\mathbf{1}')$. This variance matrix is singular and is chosen such that $\mathbf{1}'\mathbf{u} = 0$.

The overall estimation procedure is to maximise the sum of the log-likelihoods of the observations taking $\mathbf{u}$ to be conditionally fixed and the likelihood of $\mathbf{u}$ as given by the frailty distribution. The first component is approximated in the region of true values $\boldsymbol{\beta}, \mathbf{u}$ by

$$\ell_1 = \text{Constant} - \frac{1}{2}\begin{bmatrix}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\mathbf{u}} - \mathbf{u}\end{bmatrix}' \mathbf{V} \begin{bmatrix}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \\ \hat{\mathbf{u}} - \mathbf{u}\end{bmatrix}.$$

The approximation is consistent with the Newton–Raphson convergence scheme, which assumes that the log-likelihood is approximately quadratic in the region of the true values.

The frailty distribution of $\mathbf{u}$ is degenerate and the logarithm of its density function is shown in the Appendix to be

$$\ell_2 = -\tfrac{1}{2}\left[(M-1)\ln 2\pi\sigma^2 + \sigma^{-2}\mathbf{u}'\mathbf{u}\right].$$

Maximising $\ell_1 + \ell_2$ gives estimators

$$\begin{bmatrix}\tilde{\boldsymbol{\beta}} \\ \tilde{\mathbf{u}}\end{bmatrix} = \begin{bmatrix}\boldsymbol{\beta}_0 \\ \mathbf{u}_0\end{bmatrix} - \mathbf{V}_1\begin{bmatrix}0 \\ \mathbf{u}_0\end{bmatrix} + \mathbf{V}_1\begin{bmatrix}\partial\ell/\partial\boldsymbol{\beta} \\ \partial\ell/\partial\mathbf{u}\end{bmatrix}_{\boldsymbol{\beta}=\boldsymbol{\beta}_0,\ \mathbf{u}=\mathbf{u}_0},$$

$$\tilde{\sigma}^2 = (M-1)^{-1}\mathbf{u}'\mathbf{u},$$

where $\mathbf{V}_1$ is the matrix orthogonal to $[\mathbf{0}', \mathbf{1}']$ that has

$$\begin{bmatrix}\partial^2\ell/\partial\boldsymbol{\beta}\,\partial\boldsymbol{\beta}' & \partial^2\ell/\partial\boldsymbol{\beta}\,\partial\mathbf{u}' \\ \partial^2\ell/\partial\mathbf{u}\,\partial\boldsymbol{\beta}' & \partial^2\ell/\partial\mathbf{u}\,\partial\mathbf{u}' + \sigma^{-2}\mathbf{I}\end{bmatrix}$$

as a g-inverse. As implied by the work of Henderson, the variance matrix for the estimators of $\boldsymbol{\beta}$ is estimated by the appropriate extraction of terms of $\mathbf{V}_1$.

4. Infections in Kidney Patients

The data given in Table 1 show the recurrence times to infection at point of insertion of the catheter for kidney patients using portable dialysis equipment. For each patient two such recurrence times are given, with each row of the table corresponding to one patient. The columns are

Column	*Variable description*
1	Patient number
2	Recurrence time
3	1 = infection occurs; 0 = censored
4, 5, . . .	Risk variable values

The risk variables are age, sex (1 = male, 2 = female), and disease type coded as 0 = GN, 1 = AN, 2 = PKD, 3 = other. The five regression variables fitted are age, sex, and presence/absence of disease types GN, AN, PKD. These variables form the regression vector $\mathbf{x}$.

The model fitted to these data is the Cox hazard model with additional frailty term, viz.,

$$h(t, \mathbf{x}) = Z\lambda(t)e^{\mathbf{x}'\boldsymbol{\beta}}.$$

The estimation procedure begins with arbitrarily selected initial values of $\boldsymbol{\beta}$, $\mathbf{u}$, and σ^2 such as zero components for $\boldsymbol{\beta}$, $\mathbf{u}$ and $\sigma^2 = 1$. The estimation equations given in Section 3 are applied iteratively with each estimate becoming the initial value of a subsequent iteration until conver-

Table 1
Recurrence data and frailty estimates

Patient number	Recurrence times	Event types	Age	Sex	Disease type	Frailty estimate
1	8, 16	1, 1	28	1	3	2.3
2	23, 13	1, 0	48	2	0	1.9
3	22, 28	1, 1	32	1	3	1.2
4	447, 318	1, 1	31-32	2	3	.5
5	30, 12	1, 1	10	1	3	1.5
6	24, 245	1, 1	16-17	2	3	1.1
7	7, 9	1, 1	51	1	0	3.0
8	511, 30	1, 1	55-56	2	0	.5
9	53, 196	1, 1	69	2	1	.7
10	15, 154	1, 1	51-52	1	0	.4
11	7, 333	1, 1	44	2	1	.6
12	141, 8	1, 0	34	2	3	1.2
13	96, 38	1, 1	35	2	1	1.4
14	149, 70	0, 0	42	2	1	.4
15	536, 25	1, 0	17	2	3	.4
16	17, 4	1, 0	60	1	1	1.1
17	185, 177	1, 1	60	2	3	.8
18	292, 114	1, 1	43-44	2	3	.8
19	22, 159	0, 0	53	2	0	.5
20	15, 108	1, 0	44	2	3	1.3
21	152, 562	1, 1	46-47	1	2	.2
22	402, 24	1, 0	30	2	3	.6
23	13, 66	1, 1	62-63	2	1	1.7
24	39, 46	1, 0	42-43	2	1	1.0
25	12, 40	1, 1	43	1	1	.7
26	113, 201	0, 1	57-58	2	1	.5
27	132, 156	1, 1	10	2	0	1.1
28	34, 30	1, 1	52	2	1	1.8
29	2, 25	1, 1	53	1	0	1.5
30	130, 26	1, 1	54	2	0	1.5
31	27, 58	1, 1	56	2	1	1.7
32	5, 43	0, 1	50-51	2	1	1.3
33	152, 30	1, 1	57	2	2	2.9
34	190, 5	1, 0	44-45	2	0	.7
35	119, 8	1, 1	22	2	3	2.2
36	54, 16	0, 0	42	2	3	.7
37	6, 78	0, 1	52	2	2	2.1
38	63, 8	1, 0	60	1	2	1.2

gence occurs. Estimates of β together with standard errors are:

Variable	*Age*	*Sex*	*GN*	*AN*	*PKD*
Regression coefficient estimate	.0063	− 1.7947	.2062	.4099	− 1.2961
Standard error	.0134	.4337	.4840	.4937	.7120

Estimates of frailties are listed in the last column of Table 1.

The estimate of σ^2 is .3821. In general, the effect of the prior distribution on frailty terms is to shrink estimates toward the origin, thereby biasing the estimate of σ^2. Nevertheless the frailty estimates given in Table 1 appear very reasonable. The only regression coefficient that is significantly large compared to its standard error is that of the sex variable, indicating a lower infection rate for female patients.

Résumé

Dans les études de survie, le risque instantané de chaque individu peut dépendre de facteurs de risque observés, mais aussi le plus souvent, d'autres facteurs inconnus ou non mesurables. Ce constituant inconnu de la fonction de risque est généralement appelé hétérogénéïté ou fragilité individuelle. Dans le cas où la durée de survie est en fait le temps d'apparition d'un certain type d'événement avec la possibilité d'observer plusieurs temps par individu, la fragilité est alors un facteur commun à de tels temps récurrents. Un modèle contenant la fragilité est proposé pour de telles mesures répétées de temps récurrents.

References

Aitchison, J. and Silvey, S. D. (1959). Maximum-likelihood estimation of parameters subject to constraints. *Annals of Mathematical Statistics* **29**, 813–828.

Clayton, D. and Cuzick, J. (1985). Multivariate generalisations of the proportional hazards model. *Journal of the Royal Statistical Society, Series A* **148**, 82–117.

Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society, Series B* **45**, 311–354.

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. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

Crowder, M. (1985). A distributional model for repeated failure time measurements. *Journal of the Royal Statistical Society, Series B* **47**, 447–452.

Gail, M. H., Santner, T. J., and Brown, C. C. (1981). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics* **36**, 255–266.

Henderson, C. R. (1963). Selection index and expected genetic advance. In *Statistical Genetics and Plant Breeding*, W. D. Hanson and H. F. Robinson (eds), 141–163. Publication 982. Washington, D.C.: National Academy of Sciences, National Research Council.

Henderson, C. R. (1973). Sire evaluation and genetic trends. In *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr Jay L. Lush*, 10–41. Champaign, Illinois: American Society of Animal Science.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423–447.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12**, 55–68.

Holt, J. D. and Prentice, R. L. (1974). Survival analysis in twin studies and matched-pairs experiments. *Biometrika* **61**, 17–30.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* **71**, 75–83.

Hougaard, P. (1986a). A class of multivariate failure time distributions. *Biometrika* **73**, 671–678.

Hougaard, P. (1986b). Survival models for heterogeneous populations derived from stable distributions. *Biometrika* **73**, 387–396.

Oakes, D. (1982). A model for association in bivariate survival data. *Journal of the Royal Statistical Society, Series B* **44**, 414–422.

Prentice, R. L., Williams, B., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika* **68**, 373–379.

Received June 1988; *revised January* 1990; *accepted May* 1990.

Appendix

The prior distribution of $\mathbf{u}$ has the singular variance matrix $\sigma^2(\mathbf{I} - \mathbf{M}^{-1}\mathbf{1}\mathbf{1}')$ consistent with $\mathbf{1}'\mathbf{u} = 0$. Letting

$$\mathbf{u}' = [U_1, U_2, \ldots, U_{M-1} \mid U_M] = [\mathbf{u}_1' \mid U_M]$$

the prior distribution of $\mathbf{u}_1$ has variance matrix $\sigma^2\mathbf{W}_1$, where

$$\mathbf{W}_1 = \mathbf{I}_{M-1} - \mathbf{M}^{-1}\mathbf{1}_{M-1}\mathbf{1}'_{M-1}, \quad \mathbf{W}_1^{-1} = \mathbf{I}_{M-1} + \mathbf{1}_{M-1}\mathbf{1}'_{M-1},$$

so that the logarithm of the prior density is

$$\ell_2 = -\tfrac{1}{2}\left[(M-1)\ln 2\pi\sigma^2 + \sigma^{-2}\mathbf{u}_1'\mathbf{W}_1^{-1}\mathbf{u}_1\right]$$

and $\mathbf{u}_1'\mathbf{W}_1^{-1}\mathbf{u}_1 = \mathbf{u}'\mathbf{u}$.