Journal of Korean
Data & Information Science Society
2005, Vol. 16, No. 4, pp. 1017~1026

Joint Modeling of Death Times and Counts Using a Random Effects Model

Hee-Chang Park[1] · John P. Klein[2]

Abstract

We consider the problem of modeling count data where the observation period is determined by the survival time of the individual under study. We assume random effects or frailty model to allow for a possible association between the death times and the counts. We assume that, given a random effect, the death times follow a Weibull distribution with a rate that depends on some covariates. For the counts, given the random effect, a Poisson process is assumed with the intensity depending on time and the covariates. A gamma model is assumed for the random effect. Maximum likelihood estimators of the model parameters are obtained. The model is applied to data set of patients with breast cancer who received a bone marrow transplant. A model for the time to death and the number of supportive transfusions a patient received is constructed and consequences of the model are examined.

keywords : bone marrow transplant, frailty model, Marquardt's method, Pascal distribution, score and likelihood ratio tests, Wald, Weibull regression model

1. Introduction

A common problem that arises in longitudinal studies is to model the effects of some explanatory factors on the number of occurrences of a given event that have occur in some time interval. For example, one may wish to model the number of transfusions given to a bone marrow transplant patient in the course of their

1) First Author : Professor, Department of Statistics, Changwon National University, Changwon, Gyeongnam, 641-773, Korea
E-mail : hcpark@changwon.ac.kr
2) Professor, Division of Biostatistics, Medical College of Wisconsin, 9701 Watertown Plank Road, Milwaukee WI 53226,. USA.
E-mail : Klein@mcw.edu

recovery, the number of admissions to the hospital of a patient with a serious illness, or the number of doses of a drug given to a patient with a heart attack in th e emergency room. We shall denote by N(t) denote the cumulative number of events that have occurred up to time t.

Data on N(t) is available only as long as the patient is under observation. Patients can be removed from the study in one of two ways. They can be removed alive or censored at either a random lost-to-follow-up time or at the end of the study, or, observation on the patient can stop due to the death of the patient. We shall let X denote the time to death and (T, δ) denote the on study time and censoring indicator ($\delta = 1$ if T=X, $\delta = 0$ if T>X). We assume that the censoring mechanism is independent of the time to death and the number of events that have occurred at a given time.

In most cases it is not reasonable to assume that N(t) and T are independent. In the bone marrow examples, patients who require more frequent blood transfusions are often having problems in maintaining their graft and as such are at higher risk for death than patients requiring fewer transfusions. Thus N(t) and X should be positively associated. We shall induce an association between the counts and the survival times by using a so called shared frailty or random effect model. Frailty models have been used to model association in survival studies by a number of authors (See, for example Clayton 1978, Nielsen et al 1992, Klein 1992). Lawless (1987) has used random effects models to model count data. Here we shall introduce a common random effect in the model for T and N(t) which induces a positive association between these two random quantities. The random effect represents the unmeasured factors that are acting simultaneously on both the number of events and the time to death. The variance of this random factor represents a measure of the strength of this positive association between the two random quantities.

In the next Section we will describe how a model using a common gamma frailty can be applied in this problem. We shall assume that, conditional on a set of potential risk factors, the time to death follows a Weibull distribution and the counts follow a Poisson process. We develop some properties of the model which help in the interpretation of the effects of the covariates on both the time to death and the number of events. In Section 3 we discuss the problem of estimating model parameters.

In Section 4 we apply these procedures to data from the Autologous Blood and Marrow Transplant Registry (ABMTR) on 701 patients with high risk breast cancer given high dose chemotherapy followed by autologous hematopoietic stem-cell support (an autologous bone marrow transplant). Patients were transplanted between 1990-1994 and followed until 1996. The median follow up time was 16.5 months with a range of 0.1 to 66 months. Three hundred and sixty-one(51.5%) of the patients died during the course of the study. Patients had median of 5 transfusions from the same donor. One hundred and forty-four

requiring no transfusions and the maximum number of transfusions was 172. Additional details of the study can be found in Antman et al(1997).

2. The Model

In this Section we present a model for the joint distribution of the time to death, X, and the number of events which occur up to time t, N(t). We let W denote a shared random effect which has a common multiplicative effect on both the rate at which death is occurring and the rate at which the events are occurring. This random effect, which is allowed to vary from person to person, is analogous to a frailty in the usual multivariate survival modeling (See Klein et al (1992)) and a model for unobserved heterogeneity in modeling count data (See Lawless (1987)). It represents common genetic, disease specific or environmental factors that were not measured on the patient which are affecting both the number of events and the time to death. Here we assume that W has a gamma distribution with a mean of 1 and a variance θ. That is,

$$f(w) = \frac{w^{1/\theta - 1}\exp[-w/\theta]}{\Gamma[1/\theta]\,\theta^{1/\theta}}, \theta \geq 0. \tag{2.1}$$

For a given patient we have two sets of covariates which are potential explanatory factors for either the time to death or for the count (or both). Suppose that there are pd covariates which are explanatory for death and pc covariates which are explanatory for the number of events. We define the pd+1 vector Z_d whose first component is equal to 1 and whose remaining pd columns are set equal to the pd explanatory covariates for death. Similarly let Zc be the pc+1 vector of risk factor for the number of events. Again for connivance the first component of Z_c is set to 1 for an intercept term. Note that the same factor can be included in both Z_d and Z_c.

Given the value of W=w (and Z_d) we assume that the time to death follows a Weibull distribution with hazard rate

$$h(t \mid Z_d) = w\alpha t^{\alpha - 1}\exp\{\beta Z_d\}, t \geq 0, \alpha \geq 0. \tag{2.2}$$

This is a standard Weibull regression model as discussed in Klein and Moeschberger (1997) with the inclusion of the random effect, w.

For the number of events we assume, given W=w (and Z_c), that N(t) follows a Poisson process with a rate

$$\lambda(t \mid Z_c) = wt^{\phi}\exp\{\gamma Z_c\}, t \geq 0, \phi \geq 0. \tag{2.3}$$

Given W we assume that N(t) and X are independent.

To study properties of this model we first need to note that N(t) is only observable as long as $t \leq X$, and that N(t)=N(X) when $X \geq t$. Thus we have, given W=w, that, with some abuse of notation,

$$\begin{aligned} P[X=t, N(s)=k \mid w] &= P[X=t, N(\min(s,t))=k \mid w] \\ &= w\alpha t^{\alpha-1}\exp\{\beta Z_d\}\exp[-wt^{\alpha}\exp\{\beta X_d\}] \\ &\quad \times \frac{[w\min(s,t)^{\phi}\exp\{\gamma Z_c\}]^k \exp[-w\min(s,t)^{\phi}\exp\{\gamma Z_c\}]}{k!}. \end{aligned} \tag{2.4}$$

Also

$$\begin{aligned} P[X>t, N(s)=k \mid w] &= \int_t^{\infty} P[X=u, N(s)=k \mid w]du \quad \text{for} \quad t \geq s \\ &= \int_t^{s} P[X=u, N(u)=k \mid w]du + P[X>s, \ N(s)=k \mid w] \quad \text{for} \quad t<s. \end{aligned} \tag{2.5}$$

To find the unconditional distribution of X and N(·) we take the expectation of (2.4) with respect W. For $\theta>0$, this yields,

$$\begin{aligned} P[X=t, N(s)=k] &= \int_0^{\infty} P[X=t, N(s)=k \mid w] \ \frac{w^{1/\theta-1}\exp[-w/\theta]}{\Gamma[1/\theta]\theta^{1/\theta}}\, dw \\ &= \frac{\Gamma(1/\theta+k+1)}{k!\Gamma(1/\theta)}[\alpha\theta t^{\alpha-1}\exp\{\beta Z_d\}] \ [\theta\min(s,t)^{\phi}\exp\{\gamma Z_c\}]^k \\ &\quad \times [1+\theta t^{\alpha}\exp\{\beta Z_d\}+\theta\min(s,t)^{\phi}\exp\{\gamma Z_c\}]^{-(1/\theta+k+1)}. \end{aligned} \tag{2.6}$$

When θ is equal to zero then W is equal to one almost surely and X and N(·) are independent Weibull and Poisson random variables, respectively.

From (2.5) we have

$$\begin{aligned} &P[X>t, N(s)=k] \\ &= \frac{\Gamma(1/\theta+k)}{k!\Gamma(1/\theta)}[\theta s^{\phi}\exp\{\gamma Z_c\}]^k \ [1+\theta t^{\alpha}\exp\{\beta Z_d\}+\theta s^{\phi}\exp\{\gamma Z_c\}]^{-(1/\theta+k)}, \end{aligned} \tag{2.7}$$

when $t \geq s$ and

$$P[X>t, N(s)=k] = \frac{\Gamma(1/\theta+k+1)}{k!\Gamma(1/\theta)}[\alpha\theta^{k+1}\exp\{\beta Z_d+k\gamma Z_c\}]$$

$$\times \int_t^s u^{k\phi+\alpha-1}\{1+\theta u^{\alpha}\exp\{\beta Z_d\}+\theta u^{\phi}\exp\{\gamma Z_c\}\}^{-(1/\theta+k+1)}du$$
$$+\frac{\Gamma(1/\theta+k)}{k!\,\Gamma(1/\theta)}[\theta s^{\phi}\exp\{\gamma Z_c\}]^k[1+\theta s^{\alpha}\exp\{\beta Z_d\}+\theta s^{\phi}\exp\{\gamma Z_c\}]^{-(1/\theta+k)}, \quad (2.8)$$

when s>t .

For this model one can show that the marginal distribution of X is a univariate Burr distribution with survival function

$$S(t)=[1+\theta t^{\alpha}\exp\{\beta Z_d\}]^{-1/\theta}. \quad (2.9)$$

For $s\le t$ the conditional distribution of N(s) given X>t follows a Pascal distribution with parameters 1/θ and q with

$$q=\frac{\theta s^{\phi}\exp\{\gamma Z_c\}}{1+\theta t^{\alpha}\exp\{\beta Z_d\}+\theta s^{\phi}\exp\{\gamma Z_c\}}. \quad (2.10)$$

That is

$$P[N(s)=k\mid X>t]=\binom{1/\theta+k-1}{k}q^k p^{1/\theta}, \quad (2.11)$$

where p=1−q. The mean number of transfusions at time s for a patient alive at time $t\ge s$ is E[N(s)|T>t]=q/(θp) and the conditional variance is V[N(s)|T>t]=q/(θ p^2).

To find the marginal distribution of N(s) we need to compute P[X≥0, N(s)=k]. From (2.8) we see that

$$P[N(s)=k]=\frac{\Gamma(1/\theta+k+1)}{k!\,\Gamma(1/\theta)}[\alpha\theta^{k+1}\exp\{\beta Z_d+k\gamma Z_c\}]$$
$$\times\int_0^s u^{k\phi+\alpha-1}\{1+\theta u^{\alpha}\exp\{\beta Z_d\}+\theta u^{\phi}\exp\{\gamma Z_c\}\}^{-(1/\theta+k+1)}du$$
$$+\frac{\Gamma(1/\theta+k)}{k!\,\Gamma(1/\theta)}[\theta s^{\phi}\exp\{\gamma Z_c\}]^k[1+\theta s^{\alpha}\exp\{\beta Z_d\}+\theta s^{\phi}\exp\{\gamma Z_c\}]^{-(1/\theta+k)}. \quad (2.12)$$

This quantity needs to be evaluated numerically.

3. Estimation of model parameters

Estimation of model parameters is based on the total number of transfusions a patient received during their period of observation. Let Ti be the on study time for the ith person and δ_i be the death indicator ($\delta_i = 1$ if dead, $\delta_i = 0$ if censored). Let Ni= Ni(Ti) be the total number of transfusion given to the patient. Note that, as opposed to Lawless (1987), we only know the total number of events an individual has experienced not the exact times at which these events have occurred. Let Z_{di} and Z_{ci} be the covariate vectors for the ith person. For individuals who die, their contribution to the likelihood is P[X=Ti, N(Ti)=Ni] which is given by (2.6). For individuals who are censored, their contribution to the likelihood is P[X>Ti, N(Ti)=Ni], which is given by (2.7). Based on a sample of size n, the log likelihood is (up to an additive constant) given by

$$\begin{aligned} LL = \sum_{i=1}^{n} \{ & \ln[\Gamma(1/\theta + N_i + \delta_i)] + \delta_i \ln[\alpha] + [\delta_i(\alpha-1) + N_i\phi]\ln[T_i] \\ & - (1/\theta + N_i + \delta_i)\ln[1 + \theta\, T_i^{\alpha}\exp\{\beta Z_{di}\} + \theta T_i^{\phi}\exp\{\gamma Z_{ci}\}] \\ & + \delta_i \beta Z_{di} + N_i \gamma Z_{ci} + (N_i + \delta_i)\ln[\theta] \} - n\ln[\Gamma(1/\theta)], \end{aligned} \tag{3.1}$$

when $\theta > 0$. For $\theta = 0$ the log likelihood is the sum of two likelihoods, $L_1(\alpha, \beta)$ and $L_2(\phi, \gamma)$, the first the usual likelihood from a Weibull regression model,

$$L_1(\alpha, \beta) = \sum_{i=1}^{n} \delta_i [(\alpha-1)\ln[T_i] + \ln[\alpha] + \beta Z_{di}] - T_i^{\alpha}\exp\{\beta Z_{di}\}$$

and the second Poisson likelihood,

$$L_2(\phi, \gamma) = \sum_{i=1}^{n} N_i [\phi \ln[T_i] + \gamma Z_{ci}] - T_i^{\phi}\exp\{\gamma Z_{ci}\}.$$

Estimates of $\theta, \alpha, \phi, \beta$ and γ are found by maximizing the log likelihood numerically. Wald, score and likelihood ratio tests for the parameters of the model can be performed using standard constructions(See Appendix B of Klein and Moeschberger(1997) for details.).

4. Example

We shall illustrate inference for this model using the data discussed in Section 1. To estimate model parameters we used Marquardt's (1963) method to numerically maximize the likelihood (3.1). This method, which a compromise between the method of steepest descent and the Newton-Raphson technique, was used since it is difficult to obtain initial estimates of the model parameters, and the number of parameters is quite large. Details of the technique are found in Appendix A of Klein and Moeschberger (1997). A FORTRAN program was written to perform the estimation procedures.

We first fit the model with all 5 factors for both the time to death and the number of transfusions. We then use a backward stepwise procedure to eliminate factors one at time to find a final model in which all factors are significant at the 5% level. The final model is reported in Table 1. Here we see that the risk factors for death are the stage of disease with patients transplanted with resistant disease or in partial remission having higher death rates; the source of the graft, with patients transplanted with both bone marrow and peripheral blood stem cells having a worst prognosis; and the lag time between diagnosis and transplant with patients being transplanted soon after diagnosis doing more poorly. For the number of transfusions, all five factors are significant. Here younger patients transplanted later than 1992 with a long time from diagnosis to transplant tend to have fewer transfusions. For stage of disease patients transplanted in complete remission or with resistant disease tend to have more transfusions, while patients given only peripheral blood stem cells tend to have fewer transfusion.

The Wald test of the hypothesis of no association between the death times and the number of transfusions (i.e. H_o: $\theta = 0$) is strongly rejected in both models. Since this is a test about a parameter on the boundary of the parameter space, the likelihood ratio test may be more appropriate. When $\theta = 0$, X and N[·] are independent, so the likelihood is the product of the typical Weibull likelihood and a Poisson likelihood. Maximum likelihood estimates can be found by maximizing these two likelihoods separately. The total log likelihood is the sum of two individual likelihoods. Table 2 shows the results of fitting the two separate models, using the covariates in the final model. Here the total likelihood is -8049.86 which yields a likelihood ratio chi-square of 7,879.20 with 1 degree of freedom, which is highly significant.

A comparison of Tables 1 and 2 shows that effect of ignoring the significant association between the death times and the number of transfusions. In the models for the death times, the factor graft source is not significant in the independence model, but is highly significant in the dependence model. In the models for the number of transfusions the two models are also different. First, the baseline

intensity, $t^{\phi}\exp\{\beta_0\}$ is decreasing in the independence model ($\phi = -0.18$), but increasing in time in the dependence model($\phi = 0.40$). Second, the sign of the regression coefficient for "partial remission" is different in the two model.

References

1. Antman, K. H. et al (1997). High-Dose Chemotherapy with Autologous Hematopoietic Stem-Cell Support for Breast Cancer in North America. *Journal of Clinical Oncology,* **15**, 1870-1879.
2. Clayton, D. G. (1978). A Model for Association in Bivariate Life Tables and Its Application in Epidemiological Studies of Familial Tendency in Chronic Disease Incidence. *Biometrika,* 5, 141-151.
3. Klein, J. P. (1992). Semiparametric Estimation of Random Effects Using The Cox Model Based on The EM Algorithm. *Biometrics*, 48, 795-806.
4. Klein, J. P. and Moeschberger, M.L. (1997). *Survival Analysis: Techniques for Censored and Truncated Data.* Springer-Verlag, New York.
5. Klein, J. P., Moeschberger, M. L., Li, Y. H. and Wang S. T. (1992). Estimating Random Effects in The Framingham Heart Study. In J. P. Klein and P. Goel, Editors, *Survival Analysis: State of The Art,* 99-120. Kluwer Academic Publishers, Boston.
6. Lawless, J.F. (1987). Regression Methods for Poisson Process Data. *JASA,* **82,** 808-815.
7. Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Non-Linear Parameters. *SIAM Journal of Applied Mathematics*, 11, 431-441.
8. Nielsen, G.G., Gill, R. D., Andersen, P. K., and Sørensen, T. I. A. (1992). A Counting Process Approach to Maximum Likelihood Estimation in Frailty Models. *Scandinavian Journal of Statistics*, 19, 25-43.

Table 1

Maximum Likelihood Estimates of Model Parameters in the Frailty Model					
Factor	Estimate	SE	D.F.	Wald X^2	p-value
Time to Death					
Intercept	-5.23	0.245			
Stage of Disease			3	125.27	<0.0001
Complete Remission	0.65	0.240	1	7.34	0.0068
Partial Remission	2.09	0.220	1	90.25	<0.0001
Resistant Disease	2.21	0.234	1	89.20	<0.0001
Graft Source			2	9.91	0.0070
Peripheral Blood Stem Cell	-0.05	0.168	1	0.09	0.7660
Marrow and Peripheral Blood	0.71	0.249	1	8.13	0.0044
Waiting Time to Transplant	-2.2×10^{-4}	6.8×10^{-5}	1	10.31	0.0013
Number of Transfusions					
Intercept	0.79	0.286			
Stage of Disease			3	57.41	<0.0001
Complete Remission	0.02	0.146	1	0.02	0.8875
Partial Remission	0.25	0.141	1	3.14	0.0762
Resistant Disease	1.05	0.152	1	47.72	<0.0001
Graft Source			2	21.82	<0.0001
Peripheral Blood Stem Cell	-0.38	0.118	1	10.37	0.0013
Marrow and Peripheral Blood	0.35	0.182	1	3.70	0.0545
Year of Transplant(90-92)	-0.32	0.088	1	13.69	0.0002
Age	0.02	0.005	1	10.58	0.0011
Waiting Time to Transplant	-7.2×10^{-5}	3.2×10^{-5}	1	5.11	0.0238
θ	1.57	0.097	1	267.46	<0.0001
α	1.31	0.057	1		
ϕ	0.40	0.050			
Log likelihood	-4110.26				

Table 2

Model Based on Independence Between the Death Times and Number of Transfusions					
Factor	Estimate	SE	D.F.	Wald X^2	p-value
Time to Death					
Intercept	-4.54	0.21			
Stage of Disease			3	122.174	<0.0001
Complete Remission	0.67	0.19	1	12.06	0.0005
Partial Remission	1.53	0.17	1	79.73	<0.0001
Resistant Disease	1.76	0.18	1	100.06	<0.0001
Graft Source			2	3.44	0.1787
Peripheral Blood Stem Cell	-0.17	0.12	1	2.17	0.1407
Marrow and Peripheral Blood	0.30	0.17	1	3.03	0.0820
Waiting Time to Transplant	-1.84×10^{-4}	5.94×10^{-5}	1	9.61	0.0019
α	1.04	0.05			
Weibull Log Likelihood	-1599.32				
Number of Transfusions					
Intercept	2.39	0.08			
Stage of Disease			3	801.60	<0.0001
Complete Remission	-0.04	0.04	1	1.42	0.2338
Partial Remission	-0.27	0.04	1	55.87	<0.0001
Resistant Disease	0.58	0.03	1	304.15	<0.0001
Graft Source			2	236.31	<0.0001
Peripheral Blood Stem Cell	-0.37	0.03	1	180.97	<0.0001
Marrow and Peripheral Blood	0.05	0.03	1	1.70	0.1924
Year of Transplant(90-92)	-0.12	0.02	1	23.73	<0.0001
Age	0.01	1.51E-03	1	83.61	<0.0001
Waiting Time to Transplant	-3.23E-05	1.20E-05	1	7.20	0.0073
ϕ	-0.18	0.01			
Poisson Log Likelihood	-6450.54				
Total Log Likelihood	-8049.86				

[received date : Aug. 2005, accepted date : Sep. 2005]