

## 공변량을 포함한 사망시간과 치료횟수의 모형화를 위한 주변환경효과모형의 적용

박희창<sup>1</sup> · 박진표<sup>2</sup>

### 요약

본 논문에서는 환자의 수명과 치료횟수의 모형화를 위해 관측가능한 공변량을 포함하는 동시에 치료횟수에 대해 관측불가능한 환경요인을 고려하기 위한 주변환경효과모형을 이용하고자 한다. 환경효과를 나타내는 분포를 감마분포로 가정하여 사망시간과 치료횟수의 결합모형을 개발하고, 모형에 포함된 모수의 추론과정을 논의하며, 개발된 모형을 Autologous Blood and Marrow Transplant Registry(ABMTR)에 등록된 환자의 자료에 적용하여 환경효과를 고려하지 않은 독립모형과 비교하여 치료횟수에 영향을 미치는 환경요인의 효과를 검토하고자 한다.

### 1. 서론

일반적으로 시계열 연구(longitudinal study)에서 자주 발생하는 문제는 설명요인의 효과를 임의의 시간구간에서 일어나는 사건의 발생횟수에 대해 모형화하는 것이다. 이러한 문제가 일어나는 경우를 구체적으로 살펴 보면, 치명적인 질환을 가진 환자의 입원횟수나 응급실에서 심장발작을 일으킨 환자의 심장약 복용횟수를 모형화한다던가, 또는 골수이식환자의 회복과정에서 그들에게 취해진 수혈횟수에 대한 모형화 등을 그 예로 들 수 있다. 이 논문에서는 환자의 사망시간과 치료 횟수에 관하여 모형화하는데 있어서 두 변수 모두에 대해 관측가능한 공변량 뿐만 아니라 치료횟수에 대한 관측불가능한 환경요인을 고려하기 위해 소위 주변환경효과모형(marginal frailty model)을 도입하고자 한다. 이러한 환경효과모형은 Vaupel 등(1979)에 의해 처음으로 소개되었는데, 그들은 일변량 생존모형에서 개인의 이질성(heterogeneity)을 설명하기 위해 frailty모형을 도입하였다. 이후로 여러 분야에서 많은 학자들에 의해 연구되었는데, 특히 Lawless(1987)는 사건발생횟수(count data)에 관한 자료를 모형화하기 위해 frailty모형을 이용하였다.

<sup>1</sup>(641-773) 경남 창원시 사림동 9, 창원대학교 통계학과 부교수

<sup>2</sup>(631-701) 경남 마산시 합포구 월영동 449 번지, 경남대학교 컴퓨터공학과 부교수

이 논문에서는 특별히 치료횟수에 영향을 미치는 환경요인의 효과가 감마분포를 따르는 것으로 가정하여 사망시간과 치료횟수의 결합모형을 개발하고자 한다. 2절에서는 이 논문에서 사용되는 기호와 기본 가정, 그리고 모형의 특성을 기술하고, 3절에서는 모형에 포함된 모수의 추론과정을 논의하며, 4절에서는 개발된 모형을 ABMTR에 등록된 환자들 중 사망위험성이 상당히 높은 유방암환자 701명의 자료에 적용하여 환경효과를 전혀 고려하지 않은 독립모형과 비교하고자 한다.

## 2. 모형

이 절에서는 사망시간  $X$ 와 시간  $t$ 까지 일어난 사건발생횟수  $N(t)$ 의 결합분포에 대한 모형화를 고려하고자 한다. 이 논문에서 고려하는 모형은 Park(1998)이 연구한 분산이 다른 개별환경효과모형(individual frailty model)에서 사망시간에 대해서는 환경효과가 유의하지 않는 경우에 존재하는 모형이다. 먼저 사망률과 사건발생률에 영향을 미치는 환경효과를  $w$ 로 표기한다. 이러한 환경효과는 각 개인마다 다른 값을 취할 수 있는데, 이는 Klein 등(1992)이 고려한 일반적인 생존분석모형에서의 frailty모형이나, Lawless(1987)가 토의한 사건발생횟수의 자료를 모형화하는 데 있어서 관측 불가능한 이질성에 대한 모형과 유사하다. 따라서 치료횟수에 영향을 미치는 환경효과는 환자에 대해서 관측이 불가능한 유전적이거나 질병 특유의, 또는 환경적인 요인을 나타낸다. 여기서  $w$ 는 평균이 1이고, 분산이  $\theta$ 인 다음의 식 (1)의 감마분포를 따르는 것으로 가정한다.

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

조사대상환자에 대해 두 집단의 공변량을 고려하는데, 이들은 사망시간과 사건발생횟수 각각에 대한 잠재적인 설명요인이거나 두 변량 모두에 대한 잠재적인 설명요인이다. 사망시간에 대해서는  $P_d$ 개의 공변량을 고려하고, 사건발생횟수에 대해서는 공변량의 수를  $P_c$ 개로 한다. 따라서  $Z_d$ 를 절편항을 나타내기 위한 첫 번째 원소값 1과  $P_d$ 개의 공변량으로 구성된 사망시간에 대한 공변량 벡터( $P_d+1$ )로 정의하고,  $Z_c$ 를 이와 유사하게 첫 번째 원소의 값 1과 사건발생횟수에 대한  $P_c$ 개의 공변량으로 구성된 공변량벡터( $P_c+1$ )로 정의한다.  $Z_d$ 와  $Z_c$ 에는 동일한 요인이 포함될 수도 있다.

공변량  $Z_d$ 가 주어졌을 때 환자의 사망시간은 다음과 같은 위험률을 갖는 Weibull 분포를 따르는 것으로 가정한다.

$$h(t|Z_d) = \alpha t^{\alpha-1} \exp(\beta Z_d), \quad t \geq 0, \alpha \geq 0 \quad (2)$$

식 (2)는 표준적인 Weibull 회귀모형이다. 사건발생횟수에 대해서는  $W$ 의 값과 공변량  $Z_c$ 가

주어졌을 때  $N(t)$ 는 다음과 같은 비율을 갖는 Poisson 과정을 따르는 것으로 가정한다.

$$\lambda(t|w, Z_c) = wt^\phi \exp(\gamma Z_c), \quad t \geq 0, \phi \geq 0 \quad (3)$$

이러한 가정하에서 개발된 모형의 성질을 규명하기 위해 먼저 환경효과  $W$ 가 주어졌을 때  $X$ 와  $N(\cdot)$ 의 결합확률을 구하여야 하는데,  $N(t)$ 는 관찰기간  $t$ 의 값이 환자의 사망시간보다 짧은 경우에만 관측가능하며, 만약  $X > t$ 인 경우에는  $N(t) = N(X)$ 가 된다는 사실하에서 다음과 같이 얻어진다.

$$\begin{aligned} P[X = t, N(s) = k|w] &= \alpha t^{\alpha-1} \exp\{\beta Z_d\} \exp[-t^\alpha \exp\{\beta Z_d\}] \\ &\times \frac{\{w(\min(s, t))^\phi \exp[\gamma Z_c]\}^k \exp[-w(\min(s, t))^\phi \exp[\gamma Z_c]]}{k!} \end{aligned} \quad (4)$$

또한  $W$ 가 주어졌을 때  $t$ 이상의 값을 취하는  $X$ 와  $N(\cdot)$ 의 결합확률은 다음과 같이 표현된다.

$$P[X \geq t, N(s) = k|w] = \begin{cases} \int_t^\infty P[X = u, N[s] = k|w] du, & \text{if } t \geq s \\ \int_t^s P[X = u, N[u] = k|w] du + P[X \geq s, N[s] = k|w], & \text{if } t < s \end{cases} \quad (5)$$

그런데  $W$ 는 관측이 불가능하므로 식 (4)를  $W$ 에 대해 기대값을 취하면 다음과 같이  $X$ 와  $N(\cdot)$ 의 결합확률을 구할 수 있다.

$$\begin{aligned} P[X = t, N(s) = k] &= \int_0^\infty P[X = t, N(s) = k|w] f(w) dw \\ &= [\alpha t^{\alpha-1} \exp(\beta Z_d)] \exp[-t^\alpha \exp(\beta Z_d)] \frac{\Gamma(1/\theta + k)}{k! \Gamma(1/\theta)} \\ &\times \left[ \frac{\theta(\min(s, t))^\phi \exp(\gamma Z_c)}{1 + \theta(\min(s, t))^\phi \exp(\gamma Z_c)} \right]^k \left[ \frac{1}{1 + \theta(\min(s, t))^\phi \exp(\gamma Z_c)} \right]^{1/\theta} \end{aligned} \quad (6)$$

이와 마찬가지로 식 (5)로부터  $t$ 이상의 값을 취하는  $X$ 와  $N(\cdot)$ 의 결합확률을 구할 수 있는데, 먼저  $t \geq s$ 인 경우에는 다음과 같다.

$$\begin{aligned} P[X \geq t, N(s) = k] &= \exp[-t^\alpha \exp(\beta Z_d)] \frac{\Gamma(1/\theta + k)}{k! \Gamma(1/\theta)} \\ &\times \left[ \frac{\theta s^\phi \exp(\gamma Z_c)}{1 + \theta s^\phi \exp(\gamma Z_c)} \right]^k \left[ \frac{1}{1 + \theta s^\phi \exp(\gamma Z_c)} \right]^{1/\theta} \end{aligned} \quad (7)$$

또한  $s > t$ 인 경우에는 다음과 같이 얻어진다.

$$\begin{aligned}
 P[X \geq t, N(s) = k] &= \frac{\Gamma(1/\theta + k)}{k! \Gamma(1/\theta)} [\alpha(1/\theta)^{1/\theta} \exp(\beta Z_d + k\gamma Z_c)] \quad (8) \\
 &\times \int_t^s u^{k\phi + \alpha - 1} \exp[-u^\alpha \exp(\beta Z_d)] [1/\theta + u^\phi \exp(\gamma Z_c)]^{-(1/\theta + k)} du \\
 &+ \exp[-s^\alpha \exp(\beta Z_d)] \frac{\Gamma(1/\theta + k)}{k! \Gamma(1/\theta)} \left[ \frac{\theta s^\phi \exp(\gamma Z_c)}{1 + \theta s^\phi \exp(\gamma Z_c)} \right]^k \left[ \frac{1}{1 + \theta s^\phi \exp(\gamma Z_c)} \right]^{1/\theta}
 \end{aligned}$$

### 3. 모수의 추정

모형에 포함된 모수의 추정은 관찰기간동안 환자들에게 행해진 수혈의 총횟수에 근거하고 있다.  $T_i$ 를  $i$ 번째 환자의 관찰기간이라 하고,  $\delta_i$ 를  $i$ 번째 환자의 사망여부를 나타내는 지시변수라 하자. 즉, 환자가 사망했을 경우에는  $\delta_i$ 의 값이 1이 되며, 환자가 중도 절단된 경우에는  $\delta_i$ 의 값이 0이 된다.  $N_i = N_i(T_i)$ 를  $i$ 번째 환자가 받은 총수혈횟수라 하자. 본 논문에서는 Lawless(1987)의 연구와는 달리 각 개인이 경험한 사건의 정확한 발생시간은 알려져 있지 않고 단지 전체 사건발생횟수만 알려진 자료에 대하여 분석하는 것이다.  $Z_{di}$ 와  $Z_{ci}$ 를  $i$ 번째 환자의 공변량벡터라고 하자. 관찰기간동안 식 (6)으로부터 얻어지는 사망한 환자들에 대한 확률값과 식 (7)로부터 얻어지는 중도 절단된 환자들에 대한 확률값을 이용하여 우도함수를 구할 수 있는데, 표본의 크기  $n$ 을 근거로 하여 대수우도함수  $LL$ 을 구하면 식 (9)의  $LL_1(\alpha, \beta)$ 와 식 (10)의  $LL_2(\phi, \theta, \gamma)$ 의 합으로 얻어진다.

$$LL_1 = (\ln[\alpha]) \sum_{i=1}^n \delta_i + (\alpha - 1) \sum_{i=1}^n \delta_i \ln[T_i] + \sum_{i=1}^n \delta_i \beta Z_{di} - \sum_{i=1}^n T_i^\alpha \exp(\beta Z_{di}) \quad (9)$$

$$\begin{aligned}
 LL_2 &= \sum_{i=1}^n \ln \Gamma(1/\theta + N_i) - \sum_{i=1}^n (1/\theta + N_i) \ln[1/\theta + T_i^\phi \exp(\gamma Z_{ci})] \quad (10) \\
 &+ \phi \sum_{i=1}^n N_i \ln[T_i] + \sum_{i=1}^n N_i \gamma Z_{ci} - \sum_{i=1}^n \ln(N_i!) - n \{ \ln[\Gamma(1/\theta)] + \frac{1}{\theta} \ln[\theta] \}
 \end{aligned}$$

여기서  $\theta > 0$ 이다.  $\theta = 0$ 인 경우에는 대수우도함수의 값이 식 (9)와 다음의 식 (11)의  $L_2(\phi, \gamma)$ 의 합으로 얻어지는데, 전자는 일반적인 Weibull 회귀모형으로부터 얻어지는 우도함수이고, 후자는 Poisson 우도함수이다.

$$L_2(\phi, \gamma) = \sum_{i=1}^n \left[ N_i \{ \phi \ln[T_i] + \gamma Z_{ci} \} - T_i^\alpha \exp[\gamma Z_{ci}] - \ln[N_i!] \right] \quad (11)$$

모수  $\theta, \alpha, \phi, \beta$ , 그리고  $\gamma$ 에 대한 추정값은 대수우도함수  $LL$ 을 수치 해석적으로 최대화함으로써 얻을 수 있다.  $LL$ 의 각 모수에 대한 1차 편미분값인 score 통계량과 관측된 정보행렬(observed information matrix)을 부록에 제시하였는데, 이는 각 모수의 2차 편도함수의 음의 값으로 구성된다. 모수의 관측된 정보행렬의 역행렬이 최우추정량의 공분산 행렬이 된다. 또한 모형에 포함된 모수에 대한 검정은 Wald검정법, 그리고 우도비 검정법 등을 이용하여 수행할 수 있다. (이에 대해 자세한 내용은 Klein과 Moeschberger(1997)을 참조하기 바람.)

#### 4. 예 제

이 절에서는 ABMTR에 등록된 환자의 자료를 이용하여 위에서 논의한 모형에 대한 추론문제를 토의하고자 한다. 이들 환자들은 1990년부터 1994년 사이에 이식수술을 받았으며, 1996년까지 추적조사가 진행되었다. 평균추적기간은 16.5개월이고 그 범위는 0.1에서 66개월이다. 환자들 중에서 361(51.1%)명이 관찰기간동안 사망하였고, 환자들은 동일한 골수 기증자로부터 평균 5회의 수혈을 받은 것으로 나타났다. 환자들 중 144명은 한 번도 수혈을 받지 않았으며, 가장 많이 수혈을 받은 환자는 그 수가 172회나 된다.

본 논문에서는 5종류의 잠재적인 위험요인을 고려하였는데 이들 중 두 요인은 연속적인 값을 취하는 공변량으로서 이식 당시의 환자의 연령(평균:44세, 범위:24.5 ~ 64.4세)과 진단에서부터 이식까지의 대기시간(평균:626일, 범위:92 ~ 4811일)이다. 다른 하나의 요인은 이식연도로서 이진(binary) 변량(90 ~ 92:44%, 93 ~ 94:56%)으로 고려하였으며, 나머지 두 요인은 범주형 변량으로서 이들 중 하나는 이식 당시 환자의 질병단계를 나타내는 것으로 primary stage 2-3 disease(PS;32.3%), metastatic disease in complete remission(CR; 18.7%), partial remission(PR; 28.0%), 그리고 resistant disease(RD; 21.1%)로 구분하였으며, 다른 하나는 이식조직원(graft source)으로, 골수(bone marrow(BM), 49.4%), 말초혈관줄기세포(peripheral blood stem cell (PBSC), 39.5%), 그리고 골수와 말초혈관줄기세포(Both)로 구분하였다. 이에 대한 보다 구체적인 내용은 Antman 등(1997)의 연구에 잘 나타나 있다.

모형에 포함된 모수의 추론을 위해 본 논문에서는 우도함수를 수치적으로 최대화하기 위해 Marquardt(1963)의 방법을 사용하였는데, 이는 steepest decent법과 Newton-Raphson기법을 절충한 수치 해석적 방법이다. 이 방법을 사용한 이유는 개발된 모형에 포함된 모수의 초기추정값을 얻기가 어려울 뿐만 아니라 추정해야 할 모수의 수가 상당히 많기 때문이다. 또한 모수의 추정과정을 위해 FORTRAN으로 프로그램을 작성하였다.

최적 모형을 구하기 위해 먼저 사망시간과 수혈횟수 양쪽 모두에 5종류의 요인 전부를 개발된 모형에 고려하였다. < 표 1 > 에 모수의 추정값과 표준오차, 그리고 모수의 값이 0이라는 가설에 대한 Wald 검정통계량과 유의 확률값을 제시하였다. 그런 후 후진단계적

과정(backward stepwise procedure)을 이용하여 유의수준 5%하에서 유의한 모든 요인을 최종 모형에 포함시키기 위해 한번에 하나씩 요인을 제거하였다. 그 결과 최종 모형은 < 표 2 >와 같이 주어졌다. 이 표로부터 알 수 있는 바와 같이 사망시간에 대한 위험요인은 질병의 단계, 이식조직원, 그리고 대기시간 등이다. 수혈횟수에 대해서는 질병의 단계와 이식조직원이 유의한 것으로 나타났다.

수혈횟수에 대한 주변환경효과가 없다는 가설  $H_0 : \theta = 0$ 에 대한 Wald 검정을 실시한 결과 초기모형과 최종 모형 모두에서 강하게 기각된다. 또한  $\theta = 0$ 일 때, 우도함수의 값은 Weibull 우도함수의 값과 Poisson 우도함수값의 곱으로 얻어진다. 각 모형에 포함되어진 모수의 최우추정량은 이들 두 우도함수를 각각 최대화함으로써 구할 수 있는데, 전체 대수우도함수는 이들 두 우도함수에 대해 대수를 취한 대수우도함수를 합한 값이다. < 표 3 >에는 이들 두 모형을 적합한 결과를 최종모형에서 포함된 공변량을 이용하여 나타내었는데, 이 표에서 보는 바와 같이 전체 우도 함수값은 -8110.62이므로, 우도비검정 통계량의 값은 8386.02이다. 따라서 이 값을 자유도가 1인  $X^2$ 분포의 임계값과 비교해 볼 때 그 귀무가설은 상당히 유의한 것으로 나타났다. 그리고 < 표 2 >와 < 표 3 >을 비교해보면 환자의 수명과 치료횟수에 작용하고 있는 환경효과의 영향을 감지할 수 있는데, 특히 이식조직원에서 Both의 값이 독립모형인 [모형 2]에서는 0.12로 유의한 반면에, 주변환경효과를 고려한 모형에서는 그 값이 -0.05로 유의하지 않은 것을 알 수 있다. 이와 같이 치료횟수에 영향을 미치고 있는 환경효과를 무시한 독립모형을 사용한 경우에는 잘못된 추론에 도달할 수 있다.

## 참고 문헌

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. Klein, J. P. and Moeschberger, M.L. (1997). *Survival analysis: techniques for censored and truncated data*, Springer-Verlag, New York.
3. 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.
4. Lawless, J.F. (1987). Regression methods for Poisson process data, *Journal of the American Statistical Association*, 82, 808-815.
5. Marquardt, D. (1963). An algorithm for least-squares estimation of non-linear parameters, *SIAM Journal of Applied Mathematics*, 11, 431-441.
6. Park, H. C. (1998). Modeling of death times and counts using a individual frailty model with different variances, submitted.
7. Vaupel, J. W., Manton, K. G. and Stallard, E. (1979). The impact of heterogeneity in

individual frailty and the dynamics of mortality, *Demography*, 16, 439-454.

< 표 1 > 초기모형에 포함된 모수의 최우추정값

구분	요인	추정값	표준오차	자유도	Wald통계량	p값	
사망시간	절편	-4.77	0.389				
	질병의 단계			3	118.32	< 0.0001	
	CR	0.67	0.193	1	11.92	0.0006	
	PR	1.53	0.173	1	78.99	< 0.0001	
	RD	1.74	0.178	1	95.45	< 0.0001	
	이식조직원			2	7.61	0.0223	
	PBSC	-0.17	0.122	1	1.95	0.1631	
	Both	0.31	0.175	1	3.05	0.0806	
	이식년도(90-92)	0.03	0.120	1	0.08	0.7832	
	연령	$5.0 \times 10^{-3}$	0.007	1	0.45	0.5001	
	대기시간	$-1.9 \times 10^{-4}$	$6.0 \times 10^{-5}$	1	9.87	0.0017	
	$\alpha$		1.03	0.047			
	Weibull 대수우도값		-1599.07				
수혈횟수	절편	2.64	0.376				
	질병의 단계			3	35.87	< 0.0001	
	CR	-0.02	0.151	1	0.02	0.8993	
	PR	-0.29	0.146	1	3.90	0.0484	
	RD	0.62	0.161	1	14.78	0.0001	
	이식조직원			2	16.14	0.0003	
	PBSC	-0.48	0.124	1	15.00	0.0001	
	이식조직원			2	16.14	0.0003	
	이식년도(90-92)	-0.11	0.115	1	0.93	0.3341	
	연령	0.01	0.007	1	2.55	0.1105	
	대기시간	$-2.3 \times 10^{-6}$	$4.0 \times 10^{-5}$	1	$3.4 \times 10^3$	0.9534	
	$\phi$		-0.22	0.058			
	$\theta$		1.69	0.098	1	297.39	< 0.0001
Poisson 대수우도값		-2316.53					
전체 대수우도값		-3915.60					

&lt; 표 2 &gt; 최종모형에 포함된 모수의 최우추정값

구분	요인	추정값	표준오차	자유도	Wald통계량	p값
사망시간	절편	-4.54	0.205			
	질병의 단계			3	122.22	< 0.0001
	CR	0.67	0.192	1	12.05	0.0005
	PR	1.53	0.172	1	79.73	< 0.0001
	RD	1.76	0.176	1	100.00	< 0.0001
	이식조직원			2	7.67	0.0216
	PBSC	-0.17	0.118	1	2.17	0.1409
	Both	0.30	0.173	1	3.03	0.0820
	대기시간	$-1.8 \times 10^{-4}$	$5.9 \times 10^{-5}$	1	9.61	0.0019
		$\alpha$	1.04	0.046		
	Weibull 대수우도값	-1601.08				
수혈횟수	절편	3.10	0.201			
	질병의 단계			3	39.25	< 0.0001
	CR	-0.04	0.149	1	0.06	0.8073
	PR	-0.31	0.141	1	4.85	0.0276
	RD	0.61	0.151	1	16.64	< 0.0001
	이식조직원			2	14.78	0.0006
	PBSC	-0.43	0.117	1	13.74	0.0002
	Both	-0.05	0.178	1	0.08	0.7794
	$\phi$	-0.24	0.057			
	$\theta$	1.72	0.101	1	290.01	< 0.0001
	Poisson 대수우도값	-2316.53				
	전체 대수우도값	-3917.61				



< 표 3 > 독립모형에 포함된 모수의 최우추정값

구분	요인	추정값	표준오차	자유도	Wald통계량	p값
사망시간	절편	-4.54	0.205			
	질병의 단계			3	122.22	< 0.0001
	CR	0.67	0.192	1	12.05	0.0005
	PR	1.53	0.172	1	79.73	< 0.0001
	RD	1.76	0.176	1	100.00	< 0.0001
	이식조직원			2	7.67	0.0216
	PBSC	-0.17	0.118	1	2.17	0.1409
	Both	0.30	0.173	1	3.03	0.0820
	대기시간	$-1.8 \times 10^{-4}$	$5.9 \times 10^{-5}$	1	9.61	0.0019
	$\alpha$	1.04	0.046			
	Weibull 대수우도값	-1660.08				
수혈횟수	절편	2.93	0.037			
	질병의 단계			3	857.13	< 0.0001
	CR	-0.08	0.035	1	5.60	0.0180
	PR	-0.32	0.034	1	89.10	< 0.0001
	RD	0.56	0.030	1	345.12	< 0.0001
	이식조직원			2	205.51	< 0.0001
	PBSC	-0.31	0.026	1	139.69	< 0.0001
	Both	0.12	0.033	1	12.93	0.0003
	$\phi$	-0.19	0.010			
	Poisson 대수우도값	-6450.54				
	전체 대수우도값	-8110.62				

## 부록

대수우도함수는 식 (9)의  $LL_1(\alpha, \beta)$  와 식 (10)의  $LL_2(\phi, \theta, \gamma)$ 의 합으로 얻어지므로 1차 편도함수인 score 통계량은 다음과 같다.

$$\frac{\partial LL_1}{\partial \alpha} = D/\alpha + S_{DLT} - \sum_{i=1}^n \ln[T_i]Y_{1i}; \quad (12)$$

$$\frac{\partial LL_1}{\partial \beta_j} = \sum_{i=1}^n \delta_i Z_{dij} - \sum_{i=1}^n Z_{dij} Y_{1i}, j = 0, 1, \dots, P_d; \quad (13)$$

$$\frac{\partial LL_2}{\partial \phi} = \sum_1^n N_i \ln[T_i] - \sum_1^n \frac{(1/\theta + N_i)(\ln[T_i])Y_{2i}}{1/\theta + Y_{2i}}; \quad (14)$$

$$\begin{aligned} \frac{\partial LL_2}{\partial \theta} &= -\frac{1}{\theta^2} \sum_1^n \Psi(1/\theta + N_i) + \frac{1}{\theta^2} \sum_1^n \ln[1/\theta + Y_{2i}] \\ &+ \frac{1}{\theta^2} \sum_1^n \frac{1/\theta + N_i}{1/\theta + Y_{2i}} + \frac{n}{\theta^2} \{\Psi(1/\theta) + \ln[\theta] - 1\}; \end{aligned} \quad (15)$$

$$\frac{\partial LL_2}{\partial \gamma_j} = \sum_1^n N_i Z_{cij} - \sum_1^n \frac{(1/\theta + N_i)Z_{cij}Y_{2i}}{1/\theta + Y_{2i}}, j = 0, 1, \dots, p_c; \quad (16)$$

여기서  $Psi(\cdot)$ 는 digamma 함수이고  $D$ 는 사망자수이며,  $S_{DLT} = \sum_{i=1}^n \delta_i \ln[T_i]$ ,  $S_{NLT} = \sum_{i=1}^n N_i \ln[T_i]$ ,  $Y_{1i} = T_i^\alpha \exp[\beta Z_{di}]$ , 그리고  $Y_{2i} = T_i^\phi \exp[\gamma Z_{ci}]$  이다.

대수우도함수의 2차 편도함수는 다음과 같다.

$$\frac{\partial^2 LL_1}{\partial \alpha^2} = -D/\alpha^2 - \sum_{i=1}^n (\ln[T_i])^2 Y_{1i}; \quad (17)$$

$$\frac{\partial^2 LL_1}{\partial \alpha \partial \beta_j} = -\sum_{i=1}^n Z_{dij} \ln[T_i] Y_{1i}, j = 0, 1, \dots, P_d; \quad (18)$$

$$\frac{\partial^2 LL_1}{\partial \beta_j \partial \beta_k} = -\sum_{i=1}^n Z_{dij} Z_{dik} Y_{1i}; \quad (19)$$

$$\frac{\partial^2 LL_2}{\partial \phi^2} = -\frac{1}{\theta} \sum_1^n \frac{(1/\theta + N_i)(\ln[T_i])^2 Y_{2i}}{(1/\theta + Y_{2i})^2}; \quad (20)$$

$$\frac{\partial^2 LL_2}{\partial \phi \partial \theta} = \frac{1}{\theta^2} \sum_1^n \frac{(\ln[T_i])Y_{2i}}{1/\theta + Y_{2i}^2} - \frac{1}{\theta^2} \sum_1^n \frac{(1/\theta + N_i)(\ln[T_i])Y_{2i}}{(1/\theta + Y_{2i})^2}; \quad (21)$$

$$\frac{\partial^2 LL_2}{\partial \phi \partial \gamma_j} = -\frac{1}{\theta} \sum_1^n \frac{(1/\theta + N_i)Z_{cij}(\ln[T_i])Y_{2i}}{(1/\theta + Y_{2i})^2}, j = 0, 1, \dots, p_c; \quad (22)$$

$$\begin{aligned}
\frac{\partial^2 LL_2}{\partial \theta^2} &= \frac{2}{\theta^3} \sum_1^n \Psi(1/\theta + N_i) + \frac{1}{\theta^4} \sum_1^n \Psi(1/\theta + N_i) - \frac{2}{\theta^3} \sum_1^n \ln[1/\theta + Y_{2i}] \\
&- \frac{2}{\theta^4} \sum_1^n \frac{1}{1/\theta + Y_{2i}} - \frac{2}{\theta^3} \sum_1^n \frac{1/\theta + N_i}{1/\theta + Y_{2i}} + \frac{1}{\theta^4} \sum_1^n \frac{1/\theta + N_i}{(1/\theta + Y_{2i})^2} \\
&- \frac{n}{\theta^3} \{2\Psi(1/\theta) + \Psi(1/\theta) + 2\ln[\theta] - 3\}; \tag{23}
\end{aligned}$$

$$\frac{\partial^2 LL_2}{\partial \theta \partial \gamma_j} = \frac{1}{\theta^2} \sum_1^n \frac{Z_{cij} Y_{2i}}{1/\theta + Y_{2i}} - \frac{1}{\theta^2} \sum_1^n \frac{(1/\theta + N_i) Z_{cij} Y_{2i}}{(1/\theta + Y_{2i})^2}, j = 0, 1, \dots, p_c; \tag{24}$$

$$\frac{\partial^2 LL_2}{\partial \gamma_j \partial \gamma_q} = -\frac{1}{\theta} \sum_{i=1}^n \frac{(1/\theta + N_i) Z_{cij} Z_{ciq} Y_{2i}}{(1/\theta + Y_{2i})^2}, j, q = 0, 1, \dots, p_c; \tag{25}$$

여기서  $\Psi(\cdot)$ 는 trigamma 함수이다.

## Joint Modeling of Death Times and Counts Considering a Marginal Frailty Model

Hee-Chang Park <sup>3</sup> · Jin-Pyo Park <sup>4</sup>

### Abstract

In this paper the problem of modeling count data where the observation period is determined by the survival time of the individual under study is considered. We assume marginal frailty model in the counts. We assume that the death times follow a Weibull distribution with a rate that depends on some covariates. For the counts, given a frailty, a Poisson process is assumed with the intensity depending on time and the covariates. A gamma model is assumed for the frailty. 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.

---

<sup>3</sup>(641-773) Associate Professor, Department of Statistics, Changwon National University

<sup>4</sup>(631-701) Associate Professor, Department of Computer Engineering, Kyungnam University