

# A Joint Frailty Model for Competing Risks Survival Data

Il Do Ha<sup>a,1</sup> · Geon-Ho Cho<sup>b</sup>

<sup>a</sup>Department of Statistics, Pukyong National University;

<sup>b</sup>Faculty of Medical Industry Convergence, Daegu Haany University

(Received October 22, 2015; Revised October 26, 2015; Accepted October 26, 2015)

---

## Abstract

Competing-risks events are often observed in a clustered clinical study such as a multi-center clinical trial. We propose a joint modelling approach via a shared frailty term for competing risks survival data from a cluster. For the inference we use the hierarchical likelihood (or h-likelihood), which avoids an intractable integration. We derive the corresponding h-likelihood procedure. The proposed method is illustrated via the analysis of a practical data set.

Keywords: competing risks models, frailty models, h-likelihood, joint models, random effects

---

## 1. 서론

최근 의학연구(예: 임상시험)에서 한 개인이 하나 이상의 사건을 경험하는 생존분석 연구가 활발히 진행되고 있으며 이는 질병진행 및 회복과정 연구에 매우 중요하다 (Putter 등, 2007; Ha 등, 2014). 한 개인이 여러 다른 종류의 사건들(several distinct events) 중 먼저 경험하는 하나의 사건으로 관측되는 생존시간(survival time 또는 event time)자료를 경쟁위험자료(competing risks data)라 부르며, 한 사건의 발생은 다른 사건의 발생을 방해한다. 예를 들어, Figure 1.1에서와 같이 어떤 암 임상연구에서 관심사건(main event 또는 event of interest)이 재발(recurrence, 상태 2)일 때, 재발 전 사망(상태 3)은 중도절단(censoring)이 아니라 경쟁사건(competing event)이 되며 이러한 경쟁사건은 관심사건의 관측을 방해한다 (Pintilie, 2006).

Figure 1.1과 같은 다중상태(multi-state)구조를 갖는 경쟁위험자료를 분석하기 위해, 기존연구에서는 관심사건과 경쟁사건은 서로 독립이라는 가정 하에서 Cox 형태의 비례위험(proportional hazards; PH)에 근거한 원인별(cause-specific) 회귀모형 (Prentice 등, 1978) 또는 부차위험(subhazard)에 기초한 회귀모형 (Fine과 Gray, 1999)을 암묵적으로 그리고 표준적으로 사용해 오고 있다. 하지만, 경쟁위험사건(competing-risks event)을 허락하는 다기관 임상연구(multi-center clinical study)와 같이 하나의 군집(cluster)내에서는 관심사건과 경쟁사건을 독립으로 가정한 모형보다는 양쪽 사건의 의존성을 고려한 모형 분석법이 타당하다고 할 수 있다 (Gorfine과 Hsu, 2011). 일반적으로 생존자료의 의존성 및

---

This work was supported by the Pukyong National University Research Fund in 2014 (C-D-2014-1260).

<sup>1</sup>Corresponding author: Department of Statistics, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Korea. E-mail: [idha1353@pknu.ac.kr](mailto:idha1353@pknu.ac.kr)

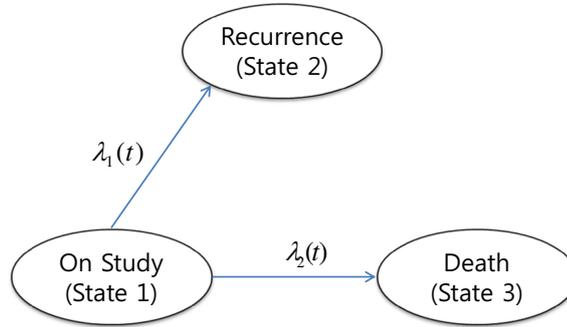


Figure 1.1. A schematic of competing risks.

이질성은 관측 안 되는 변량효과(unobserved random effect)인 프레일티를 이용하여 모형화 할 수 있다 (Hougaard, 2000; Ha 등, 2001; Duchateau와 Janssen, 2008). 본 논문에서는 군집(예: 임상시험 기관)내 환자들 간 경쟁위험 생존자료의 상관성을 설명하는 원인별 경쟁위험모형에 하나의 공통(shared) 프레일티를 허락하는 결합 프레일티모형을 고려한다.

이러한 결합모형은 프레일티 항을 포함하기 때문에 그 추론을 위해서 전통적인 주변 가능도(marginal likelihood)를 적용할 경우 매우 어려운 적분을 해야 하는 문제점에 직면할 수 있다. 본 논문에서는 다양한 변량효과 모형에 대해 적분을 피하면서 효율적인 추론결과를 제공해 주는 다단계 가능도(hierarchical likelihood 또는 h-likelihood; Lee와 Nelder, 1996; Lee 등, 2006)를 이용하여 결합 경쟁위험 모형접근법을 제안한다. 특히 다단계 가능도 방법의 사용으로 인해 기존의 유용한 다단계 가능도 방법들(예: 모형선택 및 기관간 이질성 검토; Ha 등, 2011, 2013)을 제안된 모형접근법에 적용할 수 있는 장점 또한 제공할 수 있다.

본 논문의 구성 체계는 다음과 같다. 2절에서는 결합 경쟁위험 프레일티모형을 소개하고, 3절에서는 다단계 가능도와 그 추정절차를 유도한다. 4절에서는 미국 NSABP(National Surgical Adjuvant Breast and Bowel Project)에서 수행된 유방암 임상시험자료의 분석을 통해 제안된 방법을 예증한다. 마지막으로 5절에서는 제안된 방법의 토론과 추후 연구에 대하여 논의하고자 한다.

## 2. 결합 경쟁위험 프레일티모형

Figure 1.1과 같은 경쟁위험구조 하에서 수반되는 생존자료를 체계적으로 분석하기 위한 결합 프레일티 모형은 다음과 같다. 비록 본 연구결과가 여러 개의 경쟁사건으로 쉽게 확장이 가능하지만, 논의의 편의상 관심사건과 경쟁사건을 각각 한 개로 설정한다.

본 논문에서는 다변량 자료의 형태인 군집화된 경쟁위험 생존자료를 고려한다. 두 사건형태 ( $k = 1, 2$ )를 고려할 때  $i$  ( $i = 1, \dots, q$ )번째 군(예: 시험기관)에 속해 있는  $j$  ( $j = 1, \dots, n_i$ )번째 환자에 대해,  $T_{ij1}$ 을 관심사건의 시간,  $T_{ij2}$ 를 경쟁사건의 시간, 그리고  $C_{ij}$ 를 중도절단시간이라 하자. 그러면 관측 가능한 확률변수는 다음과 같이 주어진다:

$$Y_{ij} = \min(T_{ij1}, T_{ij2}, C_{ij}), \quad \delta_{ijk} = I(Y_{ij} = T_{ijk}), \quad k = 1, 2.$$

$i$ 번째 군에 속해있는 관심사건과 경쟁사건에 대한 공통된 로그 프레일티  $v_i$ 가 주어질 때,  $T_{ij1}$ 과  $T_{ij2}$ 의 조건부 위험률은  $v_i$ 를 통해 다음과 같은 결합 위험률모형의 형태로 표현된다. 즉, 원인 1 (또는 Type

1)과 원인 2 (또는 Type 2)에 의한 각각의 사건시간  $T_{ij1}$ 과  $T_{ij2}$ 의 조건부 위험률은 다음과 같은 Cox (1972)의 비례위험모형에 프레이리티를 추가한 형태로 표현된다.

$$\lambda_{ij1}(t|v_i) = \lambda_{01}(t) \exp \left( x_{ij}^T \beta_1 + v_i \right), \tag{2.1}$$

$$\lambda_{ij2}(t|v_i) = \lambda_{02}(t) \exp \left( x_{ij}^T \beta_2 + \rho v_i \right). \tag{2.2}$$

여기서  $\lambda_{01}(\cdot)$ 과  $\lambda_{02}(\cdot)$ 는 각 사건의 미지의 기저(unknown baseline)위험함수,  $\beta_1$ 과  $\beta_2$ 는  $T_{ij1}$ 과  $T_{ij2}$ 의 공통된 공변량의 벡터  $x_{ij} = (x_{ij1}, \dots, x_{ijp})^T$ 에 대응하는 회귀모수들이며,  $v_i \sim N(0, \sigma^2)$ 로서 정규분포를 따른다고 가정한다. 특히  $\rho$ 는  $\lambda_{ij1}$ 과  $\lambda_{ij2}$ 에 서로 다른 영향을 미치는 의존성을 나타내는 하나의 모수로서,  $\rho > 0$ 이면 높은 프레이리티를 갖는 개인은 빨리 경쟁사건(competing event)을 경험하고  $\rho < 0$ 이면 그 반대를 의미한다. 위의 모형은 실패시간과 정보적 중도절단시간(informative censoring time)간의 의존성을 설명하기 위해 Huang과 Wolfe (2002)에 의해 제안된 공통된 결합프레이리티 모형으로 볼 수 있다. 전통적인 원인별 경쟁위험 회귀분석에서는 원인 2로 부터 실패한 환자들은 원인 1에 대해서는 중도절단된 것으로 취급한다. 이것은 두 사건 시간들간의 의존성을 무시하는 결과로 잘못된 추론결과를 줄 수 있다 (Gorfine과 Hsu, 2011).

### 3. 다단계 가능도와 추정절차

결합모형 (2.1)과 (2.2)에 대한 다단계 가능도 (Lee와 Nelder, 1996; Ha 등, 2001)는 다음과 같이 정의된다:

$$h = h(\beta, \lambda_0, v, \rho, \sigma^2) = \sum_{ijk} \ell_{ijk} + \sum_i \ell_i, \tag{3.1}$$

여기서

$$\ell_{ijk} = \ell_{ijk}(\beta_k, \lambda_{0k}, \rho; y_{ij}, \delta_{ijk}|v_i) = \delta_{ijk}(\log \lambda_{0k}(y_{ij}) + \eta_{ijk}) - \Lambda_{0k}(y_{ij}) \exp(\eta_{ijk})$$

는  $v_i$ 가 주어질 때 관측되는 확률변수  $(Y_{ij}, \delta_{ijk})$ 의 조건부 로그 가능도(conditional log-likelihood)이고,

$$\ell_i = \ell_i(\sigma^2; v_i) = -\frac{\log(2\pi\sigma^2)}{2} - \frac{v_i^2}{2\sigma^2}$$

는 프레이리티  $v_i$ 의 로그 가능도이다. 여기서  $\eta_{ij1} = x_{ij}^T \beta_1 + v_i$ 이고  $\eta_{ij2} = x_{ij}^T \beta_2 + \rho v_i$ 이다.

하지만 식 (3.1)에서  $\lambda_{01}(t)$ 과  $\lambda_{02}(t)$ 의 함수 형태를 전혀 모르기 때문에 관심 회귀모수  $\beta_1$ 과  $\beta_2$ 의 추론을 위해 이러한 기저위험함수를 제거한 단면 다단계 가능도(profile h-likelihood)  $h^* = h|_{\lambda_{0k} = \hat{\lambda}_{0k}}$ 를 사용한다 (Ha 등, 2001). 여기서  $\hat{\lambda}_{0k}$ 는  $\partial h / \partial \lambda_{0k} = 0$ 의 해로서 사건  $k$ 에 대한 비모수적 기저위험함수의 추정량이다. 특히  $h^*$ 는 벌점화 편 가능도(penalized partial likelihood; Therneau와 Grambsch, 2000)  $h_p$ 와 동치관계이다 (Ha 등, 2001). 따라서 본 논문에서는  $h_p$ 를 편 다단계 가능도(partial h-likelihood)라 부른다. 그러면  $h_p$ 를 이용한 추정절차는 다음과 같이 유도된다. 먼저  $(\beta_1, \beta_2, v)$ 의 점수 방정식(score equation)을 구하면 다음과 같다.

$$\begin{aligned} \frac{\partial h_p}{\partial \beta_1} &= X^T (\delta_1 - \hat{\mu}_1), \\ \frac{\partial h_p}{\partial \beta_2} &= X^T (\delta_2 - \hat{\mu}_2), \\ \frac{\partial h_p}{\partial v} &= Z^T (\delta_1 - \hat{\mu}_1) + \rho Z^T (\delta_2 - \hat{\mu}_2) - \frac{v}{\sigma^2}, \end{aligned}$$

여기서  $X$ 와  $Z$ 는 각각 회귀모수  $\beta$ 와 로그-프레일티  $v$ 에 대한 모형행렬이며,  $\hat{\mu}_1 = \hat{\Lambda}_{01} \exp(\eta_1)$ ,  $\hat{\mu}_2 = \hat{\Lambda}_{02} \exp(\eta_2)$ ,  $\hat{\Lambda}_{0k}(t) = \sum_{r: y_{(kr)} \leq t} \hat{\lambda}_{0kr}$  ( $k = 1, 2$ ),  $\eta_1 = X\beta_1 + Zv$ 이고  $\eta_2 = X\beta_2 + \rho Zv$ 이다. 또한  $y_{(kr)}$  ( $r = 1, 2, \dots, D_k$ )은  $Y_{ij}$ 들 중 순서화된 Type  $K$ 의 사건시간들이며,  $\lambda_{0k} = \lambda_{0k}(y_{(kr)})$ 이다. 위의 점수 방정식은 Ha와 Lee (2003)의 다단계 가능도 방법에 의해 아래와 같은 하나의 편리한 추정방정식으로 유도될 수 있다.

$$\begin{pmatrix} X^T W_1^* X & 0 & X^T W_1^* Z \\ 0 & X^T W_2^* X & X^T (\rho W_2^*) Z \\ Z^T W_1^* X & Z^T (\rho W_2^*) X & Z^T (W_1^* + \rho^2 W_2^*) Z + U \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{v} \end{pmatrix} = \begin{pmatrix} X^T w_1^* \\ X^T w_2^* \\ Z^T w_1^* + \rho Z^T w_2^* \end{pmatrix}, \quad (3.2)$$

여기서  $W_k^* = -\partial^2 h_p / \partial \eta_k \partial \eta_k^T$  ( $k = 1, 2$ ),  $U = -\partial^2 \ell_2 / \partial v^2 = \text{diag}(1/\sigma^2)$ 는  $q \times q$  대각행렬(diagonal matrix)이고  $w_k^* = W_k^* \eta_k + (\delta_k - \mu_k)$  ( $k = 1, 2$ )이다.

다음으로, 프레일티 모수(즉 산포모수)  $\theta = (\rho, \sigma^2)^T$ 의 추론을 위해 다음의 조정된(adjusted) 편 다단계 가능도  $p_{\beta, v}(h_p)$ 를 사용한다 (Ha와 Lee, 2003).

$$p_{\beta, v}(h_p) = \hat{h}_p - \frac{1}{2} \log \left( \det \left( \hat{H} \right) \right) + \frac{(2p + q)}{2} \log(2\pi), \quad (3.3)$$

여기서  $\hat{h}_p = h_p(\hat{\beta}(\theta), \hat{v}(\theta), \theta)$ 은  $\beta$ 와  $v$ 에 대한 단면 가능도이고  $\hat{H} = H(\hat{\beta}(\theta), \hat{v}(\theta), \theta)$ 은  $\beta$ 와  $v$ 에 대한 관측된 정보행렬(observed information matrix)이다. 그러면  $\theta$ 에 대한 추정량은 다음의 추정방정식에 의해 얻어진다:

$$\frac{\partial p_{\beta, v}(h_p)}{\partial \theta} = 0, \quad (3.4)$$

여기서  $p_{\beta, v}(h_p)$ 의  $m$  ( $m = 1, 2$ )번째 요소는 다음과 같이 표현된다:

$$\frac{\partial p_{\beta, v}(h_p)}{\partial \theta_m} = \frac{\partial \hat{h}_p}{\partial \theta_m} - \frac{1}{2} \text{trace} \left( \hat{H}^{-1} \frac{\partial \hat{H}}{\partial \theta_m} \right).$$

다만,  $H$ 는 관측된 정보행렬로서 식 (3.2)의 좌측의 정방행렬이며  $\hat{H} = H(\hat{\beta}(\theta), \hat{v}(\theta), \theta)$ 이다. 사실 이러한 추정량은 하나의 제한최대가능도(restricted maximum likelihood) 추정량의 한 확장이다. 그러면 관심모수  $(\beta_1, \beta_2, \rho, \sigma^2)$  및 프레일티  $v$ 의 추정치는 적절한 초기추정치 및 뉴턴-랩슨방법을 이용하여, 두 개의 추정방정식 (3.2)와 (3.4)를 반복적으로 적용함으로써 얻어진다 (Ha 등, 2011).

## 4. 실제 자료분석

### 4.1. 자료의 설명

2절에서 제시된 결합 경쟁위험 프레일티모형의 예증을 위해 미국 NSABP에서 수행된 유방암 환자에 대한 제 3상 무작위대조 이중맹검 다기관임상시험(randomized-controlled double-blind multi-center clinical trial)으로부터 수집된 경쟁위험 생존자료를 고려한다 (Fisher 등, 1996). 이 시험은 유방암 수술 후 타목시펜(Tamoxifen) 치료약이 가짜약(Placebo)보다 재발 위험률을 유의하게 감소시켜 주는지 조사하는 것이 목적이다. 분석의 목적을 위해 종양의 크기가 2.5cm 이상인 고위험 여성 유방암환자(즉 167개의 시험기관으로부터 모집된 731명)를 대상으로 하였다. 여기서, 이러한 경쟁위험구조하에서 환자의 다중사건(multiple events)도 허용하였다.

주 평가변수(primary variable)는 사건시간으로서 수술 후 국소 또는 구역 재발(local or regional recurrence)까지의 시간이다(단위: year). 반응 결과값(outcomes)은 2가지 형태의 사건시간과 중도절단시간으로 구성되며 아래와 같이 요약된다.

**Table 4.1.** Estimates of parameters in independent (Indep) and joint (Joint) frailty models and model selection for breast cancer data

Model	Event	Age (SE)	Trt (SE)	$\hat{\theta}$	AIC
Indep	Type I	-0.017 (0.009)	-0.633 (0.199)	$\hat{\sigma}_1^2 = 0.12$	cAIC = 6979.6
	Type II	0.014 (0.005)	-0.123 (0.094)	$\hat{\sigma}_2^2 = 0.001$	rAIC = 6996.4
Joint	Type I	-0.016 (0.010)	-0.738 (0.233)	$\hat{\sigma}^2 = 1.74$	cAIC = 6712.7
	Type II	0.013 (0.007)	-0.259 (0.149)	$\hat{\rho} = 1.10$	rAIC = 6855.6

- (i) 유형 I(Type I): 국소 또는 구역 재발로서 그 비율은 13.5%이다.
- (ii) 유형 II(Type II): 원격재발(distance recurrence), 새로운 주요한 암 발생 또는 사망로서 그 비율은 53.8%이다.
- (iii) 중도절단(censoring): 최종 추적 때 까지 어떤 사건의 발생 없이 생존한 경우로서 그 비율은 32.7%이다.

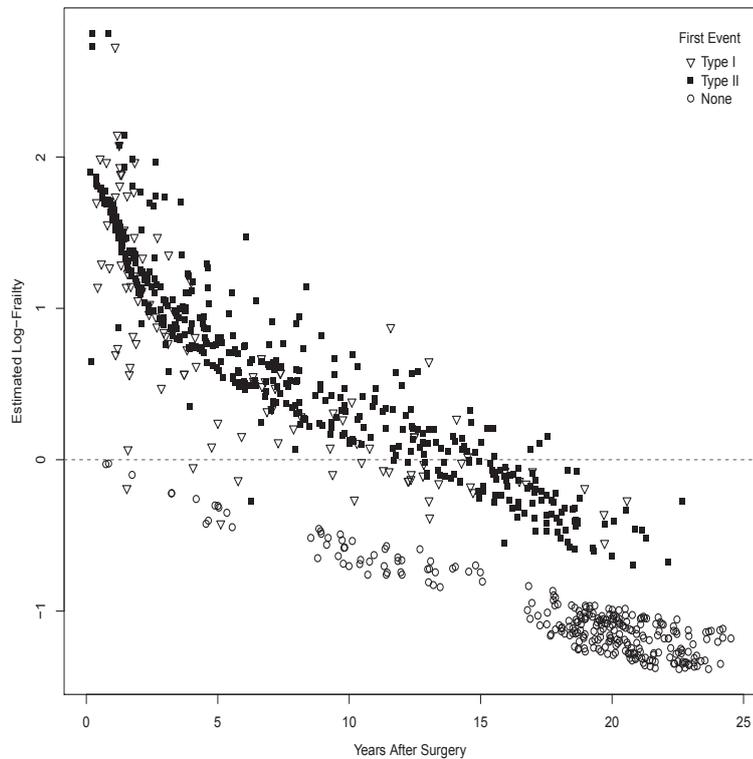
본 논문의 분석에서 관심 있는 공변량은 다음과 같다.

- (i) 나이(Age): 연속형 변수로 25세에서 75세의 범위를 가지며 평균나이는 54세이다.
- (ii) 치료법(Trt): 이진형 변수로 타목시펜 (360명 배정)은 “1”이고 가짜약(371명 배정)은 “0”으로 코딩하였다.

#### 4.2. 모형적합 및 결과분석

3절에서 제안한 다단계 가능도 방법을 이용하여 위의 두 공변량을 가진 결합 프레일티모형(“Joint”로 표시)을 적합하였다. 여기서 프레일티는 다중사건을 허용하면서 군집의 역할을 하는 환자에 대한 공통된(shared) 변량효과를 지칭한다. 비교를 위해 유형 I과 유형 II에 각각의 프레일티항( $v_1$ 과  $v_2$ )을 갖는 독립적인 프레일티 모형(“Indep”로 표시)을 추가하였다. 여기서  $v_1 \sim N(0, \sigma_1^2)$ 와  $v_2 \sim N(0, \sigma_2^2)$ 은 서로 독립이다. 두 모형의 적합 결과는 Table 4.1과 같다. 먼저 “Joint”모형에서 의존성 모수의 추정치  $\hat{\rho} = 1.10$ 로 양수를 나타내므로 높은 프레일티를 갖는 유방암 환자는 빨리 경쟁사건을 경험할 가능성이 높음을 알 수 있다. 로그-프레일티의 분산인  $\hat{\sigma}^2 = 1.74$ 로서 환자 간 사건시간에 있어서 이질성이 다소 높음을 시사한다 (Figure 4.1). 회귀모수 추정결과를 살펴보면 다음과 같다. 유형 I의 경우 나이 (Est. = -0.016, SE = 0.010)가 증가할수록 재발률이 감소하지만 유의하지는 않다. 하지만 주 관심대상인 치료법(Trt)인 경우 타목시펜 (Est. = -0.738, SE = 0.233)이 가짜약보다 재발률을 5% 유의수준에서 유의하게 감소시킴을 알 수 있다. 한편 유형 II의 경우 나이의 회귀 추정치는 유형 I과 달리 양수 (Est. = 0.013, SE = 0.007)이지만 유의하지는 않으며, 치료법 또한 유의하지는 않다 (Est. = -0.259, SE = 0.149). 다음으로, “Indep”모형에서의 회귀추정치들은 전반적으로 “Joint”모형과 유사한 경향을 보이지만, 두 사건의 의존성을 무시함으로 인해 치료법 효과의 절대값이 상대적으로 작게 추정되고 있음을 알 수 있다. 더욱이 프레일티 모형에 대한 모형선택기준 (Ha 등, 2007, 2012)에 의하면 cAIC(= conditional AIC)와 rAIC(= restricted AIC) 모두 “Joint”모형에서 두 AIC값이 “Indep”모형보다 훨씬 작기 때문에, “Joint”모형을 선택한다.

나아가, “Joint”모형을 대상으로 하여 다단계 가능도 방법의 한 장점인 프레일티 자체의 추정을 통해 자료 간 이질성을 검토하는 방법을 제시하고자 한다 (Ha 등, 2013). 앞서서도 언급한 바와 같이 “Joint”모형에서 로그-프레일티 분산의 추정치는  $\hat{\sigma}^2 = 1.74$ 로서 매우 큰 값을 가진다. 이는 환자들 간 생존시간에 이질성이 심하다는 것을 의미한다. 따라서 이러한 결과는 프레일티의 효과를 잘 반영하



**Figure 4.1.** Estimated cause-specific frailties vs the first observed event time for each subject in joint model.

로, 시간에 따른 프레일티 효과를 시각화 해 주는 Figure 4.1은 이와 같은 프레일티 효과의 설명에 매우 유용하다. Figure 4.1에 의하면, 초기에 사건(즉 재발, 사망 등)을 경험한 환자는 높은 프레일티 값(매우 연약함)을 주지만, 나중에 사건을 경험한 환자는 낮은 프레일티 값(덜 연약함)을 준다. 따라서 시간이 경과함에 따라 프레일티 값이 감소함을 알 수 있다. 특히 중도 절단된 환자 (Figure 4.1에서 “None”으로 표시)는 사건을 경험하지 않아서 전반적으로 시간에 관계없이 상대적으로 낮은 프레일티 값(즉 0보다 작은 로그-프레일티의 값)을 줌을 알 수 있기 때문에, 사건을 경험한 환자보다 경험하지 않은 환자가 훨씬 덜 위험, 즉 덜 연약함(less frail)함을 알 수 있다. 또한 이 자료에 의하면 사건의 경험 유무에 따라 이질성이 뚜렷이 구분됨을 알 수 있다.

## 5. 토론

본 논문에서는 군집화된 경쟁위험 생존자료를 분석하기 위한 결합 프레일티 모형접근법을 제안하였다. 그 모형 추론을 위해 다단계 가능도의 추론절차를 유도하였고, 이를 실제 유방암자료에 적용하여 제안된 방법의 유용성을 예증하였다. 특히 Table 4.1에서 제시한 바와 같이 두 사건의 의존성을 무시한 독립모형(“Indep”)을 적합시키면 회귀모수에 대해 편이된 추론결과를 줄 수 있음을 결합모형과 비교함으로써 이러한 사실을 보였다. 다단계 가능도는 주변 가능도와 달리 프레일티 자체를 바로 추정할 수 있으므로 인해 Figure 4.1에서 제시한 바와 같이 자료간 이질성을 도식화 해주는 유용한 정보를 제공하는 장점이 있다.

본 논문에서 사용된 자료는 두 가지 사건형태를 제시하고 있지만 제안된 방법은 세 가지 사건 이상으로 확장될 수 있다. 나아가 경쟁위험구조 하에서 한 개인에게 두 번 이상의 사건발생을 허용하면 통상적으로 다중상태 모형(multi-state models)이 요구된다. 예를 들어, 준-경쟁위험모형(semi-competing risks models; Fine 등, 2001)으로의 확장은 흥미있는 추후 연구과제가 될 것이다.

## References

- Cox, D. R. (1972). Regression models and life tables(with discussion), *Journal of the Royal Statistical Society B*, **34**, 187–220.
- Duchateau, L. and Janssen, P. (2008). *The Frailty Models*, Springer, New York.
- Fine, J. P. and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk, *Journal of the American Statistical Association*, **94**, 496–509.
- Fine, J. P., Jiang, H. and Chappell, R. (2001). On semi-competing risks data, *Biometrika*, **88**, 907–919.
- Fisher, B., Dignam, J., Bryant, J., DeCillis, A., Wickerham, D. L., Wolmark, N., Costantino, J., Redmond, C., Fisher, E. R., Bowman, D. M., Deschenes, L., Dimitrov, N. V., Margolese, R. G., Robidoux, A., Shibata, H., Terz, J., Paterson, A. H. G., Feldman, M. I., Farrar, W., Evans, J. and Lickley, H. L. (1996). Five versus more than five years of tamoxifen therapy for breast cancer patients with negative lymph nodes and estrogen receptor-positive tumors, *Journal of the National Cancer Institute*, **88**, 1529–1542.
- Gorfine, M. and Hsu, L. (2011). Frailty-based competing risks model for multivariate survival data, *Biometrics*, **67**, 415–426.
- Ha, I. D., Christian, N. J., Jeong, J.-H., Park, J. and Lee, Y. (2014). Analysis of clustered competing risks data using subdistribution hazard models with multivariate frailties, *Statistical Methods in Medical Research*, Published online: 11/March/2014.
- Ha, I. D. and Lee, Y. (2003). Estimating frailty models via Poisson hierarchical generalized linear models, *Computational and Graphical Statistics*, **12**, 663–681.
- Ha, I. D., Lee, Y. and MacKenzie, G. (2007). Model selection for multi-component frailty models, *Statistics in Medicine*, **22**, 4790–4807.
- Ha, I. D., Lee, Y. and Song, J.-K. (2001). Hierarchical likelihood approach for frailty models, *Biometrika*, **88**, 233–243.
- Ha, I. D., Noh, M. and Lee, Y. (2012). frailtyHL: A package for fitting frailty models with h-likelihood, *The R Journal*, **4**, 307–320.
- Ha, I. D., Sylvester, R., Legrand, C. and MacKenzie, G. (2011). Frailty modelling for survival data from multi-centre clinical trials, *Statistics in Medicine*, **30**, 28–37.
- Ha, I. D., Vaida, F. and Lee, Y. (2013). Interval estimation of random effects in proportional hazards models with frailties, *Statistical Methods in Medical Research*, Published online: 29/January/2013.
- Hougaard, P. (2000). *Analysis of Multivariate Survival Data*, Springer, New York.
- Huang, X. and Wolfe, R. (2002). A frailty model for informative censoring, *Biometrics*, **58**, 510–520.
- Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion), *Journal of the Royal Statistical Society B*, **58**, 619–678.
- Lee, Y., Nelder, J. A. and Pawitan, Y. (2006). *Generalised Linear Models with Random Effects: Unified Analysis via h-Likelihood*, Chapman and Hall, London.
- Pintilie, M. (2006). *Competing Risks: A Practical Perspective*, Wiley, Chichester.
- Prentice, R., Kalbfleisch, J. D., Peterson, A. V., Flournoy, N., Farewell, V. T. and Breslow, N. E. (1978). The analysis of failure times in the presence of competing risks, *Biometrics*, **34**, 541–554.
- Putter, H., Fiocco, M. and Geskus, R. B. (2007). Tutorial in biostatistics: Competing risks and multi-state models, *Statistics in Medicine*, **26**, 2389–2430.
- Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*, Springer, New York.

# 경쟁위험 생존자료에 대한 결합 프레일티모형

하일도<sup>a,1</sup> · 조건호<sup>b</sup>

<sup>a</sup>부경대학교 통계학과, <sup>b</sup>대구한의대학교 의료산업융합학부

(2015년 10월 22일 접수, 2015년 10월 26일 수정, 2015년 10월 26일 채택)

---

## 요약

경쟁위험사건들은 다기관 임상시험과 같은 군집화된 임상연구에서 자주 관측되어진다. 본 논문에서는 하나의 군집으로 부터 얻어지는 경쟁위험 생존자료에 대해 공통 프레일티를 허락하는 결합 프레일티모형 접근법을 제안한다. 추론을 위해 어려운 적분 자체를 피하는 다단계 가능도를 사용하여, 대응하는 추론절차를 유도한다. 또한 실제자료 분석을 통해 제안된 방법을 예증한다.

주요용어: 경쟁위험모형, 프레일티 모형, 다단계 가능도, 결합모형, 변량효과

---

이 논문은 2014학년도 부경대학교의 지원을 받아 수행된 연구임 (C-D-2014-1260).

<sup>1</sup>교신저자: (48513) 부산광역시 남구 용소로 45, 부경대학교 통계학과. E-mail: idha1353@pknu.ac.kr