信賴性應用硏究
제2권, 제2호, pp. 121-134, 2002

비례위험모형에서 정보적 중도절단의 효과

정대현*, 홍승만**, 원동유***

*충북 청주시 흥덕구 개신동 산 48 충북대학교 자연과학대학 통계학과 교수
**충남 연기군 조치원읍 서창동 208 고려대학교 자연과학대학 정보통계학과 교수
***대전광역시 동구 자양동 17-2번지 우송대학교 컴퓨터과학과 교수

Effects of Informative Censoring in the Proportional Hazards Model

Daehyun Chung*, Seungman Hong**, DongYu Won***

*Dept. of Statistics, Chungbuk National University, Cheongju, Chungbuk
**Dept. of Informational Statistics, Korea University, Chochiwon, Chungnam
***Dept. of Computer Science, Woosong University, Daejeon

Abstract

This paper concerns informative censoring and some of the difficulties it creates in analysis of survival data. For analyzing censored data, misclassification of informative censoring into random censoring is often unavoidable. It is worthwhile to investigate the impact of neglecting informative censoring on the estimation of the parameters of the proportional hazards model. The proposed model includes a primary failure which can be censored informatively or randomly and a followup failure which may be censored randomly. Simulation shows that the loss is about 30% with regard to the confidence interval if we neglect the informative censoring.

Key Words : proportional hazard model, random censored, informative censored, martingales, counting process, partial likelihood

1. 서 론

생존분석에서 대표적인 중도절단의 형태는 정보적 중도절단과 임의 중도절단이며 특별히 임상연구에서는 정보적 중도절단이 자주 발생한다. 대부분 연구에서는 정보적 중도절단을 임의의 중도절단으로 처리하고 있다.

일반적으로 생존함수의 추론에서는 정보적 중도절단을 고려하기 위하여 정보적 중도절단을 실패사건에 정보를 제공하는 또는 실패사건의 형태로 간주하여 경쟁위험모형을 적용하여 분석한다.

정보적 중도절단 자료를 고려한 모수적 방법으로는 Marshall과 Olkin(1967)이 T_1(실패사건 시간)과 T_2(중도절단 시간)이 종속적이며 (T_1, T_2)가 이변량 지수분포를 따를 때 생존함수를 추정하였다. Nádas(1971)는 이변량 정규분포에서 조생존함수가 순생존함수를 결정한다는 것을 증명하였고 다변량 정규분포임을 가정한 경우 모수의 추정에 관한 문제는 Smith와 Helms(1995)에 의하여 다루어졌다.

대표적인 준모수적 접근 방법을 제시한 Klein, Keiding과 Kamby(1989)는 서로 다른 두 지점에서 발생하는 두 종류의 실패사건을 모형화하기 위하여 Marshall과 Olkin (1967)의 준모수 모형으로부터 주변위험 발생에 초점을 두어 비례위험 모형을 제안하였다.

실패사건과 중도절단자료의 독립성을 가정하는 것이 무리일 경우 중도절단 자료를 무시한다면 과대 또는 과소 생존함수의 추정량이 얻어진다. (Lagakos (1979)).

대부분 연구에서는 정보적 중도절단을 임의의 중도절단으로 처리하고 있다. 예를 들면 소아백혈병 환자를 대상으로 화학요법과 골수이식요법의 효과에 차이가 있는지를 알아보는 연구에서 골수이식요법에 배정된 환자들만을 고려해보자. 골수이식을 받는 것을 실패 사건으로 볼 때 골수이식을 기다리는 시간은 환자의 나이, 성별, 인종 등에 영향을 받는다고 볼 수 있다. 골수이식을 기다리고 있는 동안 백혈병이 재발되든지 또는 사망하여 골수 이식을 받을 수 없게 될 수도 있으며 또한 여러 가지 다른 이유로 골수이식을 받기 전에 연구에서 제외될 수도 있다. 첫 번째 경우는 백혈병의 병세가 악화되어 골수이식을 받을 수 없는 것으로 볼 수 있으므로 골수이식을 받을 때까지의 기간은 재발 또는 사망으로 인하여 정보적 중도절단이 되었다고 볼 수 있다. 그러나 현실적으로 충분한 추적이 이루어지지 않아 정보적 중도절단을 임의적 중도 절단으로 분류하여 자료를 분석하는 우를 범하기 쉽다

본 연구에서는 정보적 중도절단을 임의적 중도 절단으로 잘 못 다루었을 때 마팅게일 편우도추정량의 변화를 여러 가지 상황에서 모의실험을 통하여 정보의 손실을 규명하여 정보적 중도절단의 중요성을 연구하고자 한다.

제2장에서는 정보적 중도절단을 또 다른 하나의 실패사건으로 간주하여 두 종류의 실패사건에 대하여 Klein, Keiding과 Kamby(1989)이 제안한 비례위험모형을 수정하여 고려해 보았고 제3장에서는 편우도 함수를 제안하였다. 제 4장에서는 모의실험을 통하여 정보적 중도절단을 무시하고 임의적 중도절단으로 잘못 분류하였을 때 추정량의 변화를 알아보았다.

2. 정보적 중도절단을 고려한 비례위험 모형

이 절에서는 경쟁위험모형에서 Kalbfleisch와 Prentice(1980)의 방법을 이용하기 위하여 정보적 중도절단을 또 다른 하나의 실패사건으로 본다. 소아 백혈병 환자의 경우 골수이식이 주요사건이며 재발되거나 사망하는 사건이 정보적 중도절단이며 골수이식 후 재발하거나 사망할 수 있으므로 환자는 두 가지 유형의 사건중 하나 또는 두 가지 모두를 경험한다고 볼 수 있다.

T_v $(v=1,2)$를 관측대상이 사건 v를 경험하는 시간이라면 $X=\min(T_1, T_2)$는 첫 번째 사건을 경험하는 시간이다. (T_1, T_2)는 연속적인 확률변수로 가정하자. 만약 V^1을 첫 번째 유형의 사건을 나타내는 변수라면 $X = T_{v^1}$과 $V^1 = v^1$은 서로 필요충분 조건이 성립한다. $\boldsymbol{Z}$를 시간에 종속되지 않는 공변량 벡터로 간주한다면 $\boldsymbol{Z} = \boldsymbol{z}$일 때 시점 t에서 전체 위험함수(overall hazard function)와 생존함수는 각각 다음과 같다.

$$\lambda(t \mid \boldsymbol{z}) = \sum_{v^1=1}^{2} \lambda_{v^1}(t \mid \boldsymbol{z}),$$

$$S(t \mid \boldsymbol{z}) = \exp\left(-\int_0^t \lambda(u \mid \boldsymbol{z})\, du\right).$$

다음으로, 두 번째 사건과 이를 경험하는 시간을 V^2와 Y라고 하자.

$V^1 = v^1$, $X = x_1$, $\boldsymbol{Z} = \boldsymbol{z}$가 주어졌을 때 $V^2 = v^2$의 시점 t에서의 조건부 위험함수를 $\lambda_{v^2 \mid v^1}(t \mid x_1, \boldsymbol{z})$라고 하면, 전체 조건부 위험함수(overall conditional hazard function)와 생존함수는 각각 다음과 같다.

$$\lambda(t \mid x_1, \boldsymbol{z}) = \sum_{v^2 \neq v^1}^{2} \lambda_{v^2 \mid v^1}(t \mid x_1, \boldsymbol{z}),$$

$$S(t \mid x_1, \boldsymbol{z}) = \exp\left(-\int_{x_1}^t \lambda(u \mid x_1, \boldsymbol{z})\, du\right).$$

공변량 벡터 $\boldsymbol{Z} = \boldsymbol{z}$일 때 첫 번째 사건 v^1이 시점 x_1과 두 번째 사건 v^2가 시점 t에서 발생한다면 결합 위험함수는 다음과 같다.

$$\lambda_{v^1, v^2}(x_1, t \mid \boldsymbol{z}) = \lambda_{v^1}(x_1 \mid \boldsymbol{z})\, \lambda_{v^2 \mid v^1}(t \mid x_1, \boldsymbol{z}) \ .$$

공변량 벡터 $\boldsymbol{Z} = \boldsymbol{z}$일 때 첫 번째 사건의 비례위험모형과 첫 번째 사건 이후에 발생하는

두 번째 사건의 조건부 비례위험모형을 다음과 같이 가정하자.

$$\lambda_{v^1}(t \mid \boldsymbol{z}) = \lambda_{0v^1} \exp(\boldsymbol{z}\boldsymbol{\beta}) ,$$

$$\lambda_{v^2 \mid v^1}(t \mid x_1 , \boldsymbol{z}) = \lambda_{0v^2} \exp(\boldsymbol{z}\boldsymbol{\beta}) h(x_1) I(t > x_1) . \tag{2.1}$$

단, λ_{0v^i} $(i = 1 , 2)$는 상수인 기저위험함수이고, $\boldsymbol{\beta}$는 회귀모수의 벡터이다. 그리고 $h(x_1)$은 첫 번째와 두 번째 실패사건 사이의 관계를 나타내는 함수이다. Klein, Keiding과 Kamby(1989)는 암세포가 처음 발견된 위치에서 다른 위치로 전이되었을 때 위험함수를 $h(t) = e^{\gamma}$로 가정하였다.

T를 골수 이식(유형 1)을 받을 때까지의 기간이고, U를 재발 또는 사망(유형 2)할 때까지의 기간이라고 한다면, 식 (2.1)은 T와 U의 관계를 나타낼 수 있는 모형으로 채택할 수 있다.

식 (2.1)의 함수 $h(x)$는 여러 가지 형태가 있을 수 있지만 가장 해석이 용이한 간단한 형태의 함수로

$$h(x_1) = \exp(\alpha x_1) I(v^1 = 1 , v^2 = 2) \tag{2.2}$$

를 고려할 수 있다.

식 (2.2)를 식 (2.1)에 대입하면 식 (2.1)은

$$\lambda_{v^1}(t \mid \boldsymbol{z}) = \lambda_{0v^1} \exp(\boldsymbol{z}\boldsymbol{\beta}) , \quad v^1 = 1 , 2 ,$$

$$\lambda_{v^2 \mid v^1}(t \mid x_1 , \boldsymbol{z}) = \lambda_{0v^2} \exp(\boldsymbol{z}\boldsymbol{\beta} + x_1 \alpha) I(t > x_1 , v^1 = 1 , v^2 = 2) .$$

이 때

$$\frac{\lambda_{v^2 \mid v^1}(t \mid x_1 + 1 , \boldsymbol{z})}{\lambda_{v^2 \mid v^1}(t \mid x_1 , \boldsymbol{z})} = e^{\alpha} ,$$

이므로 α는 시점 x_1과 시점 $(x_1 + 1)$에서 골수 이식을 받은 환자들의 시점 t에서 재발하거나 사망한 로그위험률을 나타낸다. 즉 α는 골수 이식을 받을 때까지의 기간이 재발 또는 사망에 미치는 효과를 나타내는 척도이다. $\alpha > 0$ 또는 $\alpha < 0$에 따라 재발하거나 사망할 위험률은 BMT 기간에 따라 증가하거나 감소한다. 마찬가지로 회귀 벡터 계수 $\boldsymbol{\beta}$는 환자의 일반적인 특성이 위험률에 미치는 정도를 나타내는 척도로 볼 수 있다.

3. 편우도 함수 (Partial Likelihood Function)

비례위험모형 식 (2.1)을 따르는 임의 중도절단 자료의 우도함수는 다음과 같이 유도된다. T와 U를 각각 사건 유형 1과 유형 2의 생존시간이라 보고 C를 어느 환자가 임의로 중도 절단된 시간이라고 하자. 환자를 관측한 결과는 1장에서 설명한 대로 4개의 범주로 나누

어 정리할 수 있다.

<표 1> 범주에 따른 관측 값

범주	관측값 특성	관측값
1	$C \le T,\ C \le U$	C
2	$U \le T,\ U \le C$	U
3	$T \le C \le U$	$T,\ C$
4	$T \le U \le C$	$T,\ U$

범주 1의 환자는 골수이식을 받기 전에 중도 절단된 경우이고 골수이식을 받기 전에 재발하거나 사망한 환자는 범주 2에 속한다. 범주 3에 속하는 환자는 골수이식을 받은 후 중도 절단된 환자이며 골수이식을 받은 후에 재발하거나 사망한 환자는 범주 4에 속하게 된다. 그리고, 만일 i번째 환자의 실패사건 시간을 T_i, 정보적 중도절단 시간을 U_i, 임의 중도절단 시간을 C_i, 공변량 벡터를 Z_i라고 하자. 그리고

$$
\begin{aligned}
X_{1i} &= \min\{T_i,\ U_i,\ C_i\}, \\
X_{2i} &= \min\{U_i,\ C_i\}, \\
\delta_i &= I\{T_i \le U_i \wedge C_i\}, \\
\xi_i &= I\{U_i \wedge C_i\}.
\end{aligned}
$$

계수과정 $N_{11i}(t)$, $N_{12i}(t)$, $Y_{1i}(t)$, $N_{2i}(t)$ 그리고 $Y_{2i}(t)$를 다음과 같이 정의하자.

$$
\begin{aligned}
N_{11i}(t) &= I\{X_{1i} \le t,\ \delta_i = 1\}, \\
N_{12i}(t) &= I\{X_{1i} \le t,\ \delta_i = 0,\ \xi_i = 1\}, \\
Y_{1i}(t) &= I\{X_{1i} \ge t\}, \\
N_{2i}(t) &= I\{X_{2i} \le t,\ \delta_i = 1,\ \xi_i = 1\}, \\
Y_{2i}(t) &= I\{X_{1i} < t,\ X_{2i} \ge t\}.
\end{aligned}
$$

주어진 계수과정 $N(\cdot)$에 대하여, $N(u) - N(u-) = 1$이면 $\Delta N(u) = 1$이라고 하고, 그 외에는 $\Delta N(u) = 0$이라고 정의하자.

Fleming과 Harrington(1991)을 참조하여 모형 식 (2.2)을 가정하여 n명의 환자로부터 구성할 수 있는 편우도 함수는 다음과 같이 표현된다.

$$PL_n(\theta) = \prod_{s=1}^{2}\prod_{i=1}^{n}\prod_{0\le u}\left[\frac{Y_{1i}(u)\exp(\boldsymbol{z}_i\boldsymbol{\beta})}{\sum_{j=1}^{n} Y_{1j}(u)\exp(\boldsymbol{z}_j\boldsymbol{\beta})}\right]^{\Delta N_{1si}(u)}$$

$$\times \prod_{i=1}^{n}\prod_{0\le u}\left[\frac{Y_{2i}(u)\exp(\boldsymbol{z}_i\boldsymbol{\beta}+x_{1i}\alpha)}{\sum_{j=1}^{n} Y_{2j}(u)\exp(\boldsymbol{z}_j\boldsymbol{\beta}+x_{1j}\alpha)}\right]^{\Delta N_{2i}(u)}$$

일반적으로 $\boldsymbol{\theta}$의 최대 편우도 추정량은 $U(\boldsymbol{\theta}) = \dfrac{\partial \ln PL_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0$ 의 해가 되므로 $\boldsymbol{\theta} = (\boldsymbol{\beta}, \alpha)$의 추정량을 구하기 위하여 위 식을 이용한다.
로그 편우도 함수가 오복함수임으로 $\boldsymbol{\theta}$의 최대 편우도 추정량의 유일성을 보장된다.

4. 정보적 중도절단을 임의적 중도절단으로 잘못 분류하였을 때 추정량의 변화 및 결론

4.1 자료의 생성

앞에서 정의된 공변량 $Z = z$가 주어졌을 때 3변량 (T, U, C)의 조건부 분포로부터 표본 $n(=50, 100)$개를 추출하여 모의 실험하는 과정을 $N(=1000)$ 번 반복 실행하였다. Z의 값 z는 $p = 0.5$인 베르누이 분포로부터 얻었다. λ_{01}, λ_{02}, β, α가 주어졌을 때 $T = x_1$, $U = u_1$, $U = u_2 \mid T = x_1$에 대한 조건부 위험률은 각각 다음과 같이 가정했다.

$$\lambda_T(x_1 \mid z) = \lambda_{01} e^{z\beta},$$

$$\lambda_U(u_1 \mid z) = \lambda_{02} e^{z\beta},$$

$$\lambda_{U\mid T}(u_2 \mid x_1, z) = \lambda_{02} e^{z\beta + x_1\alpha} I(u_2 > x_1).$$

T와 U의 결합밀도함수는 아래와 같이 나타낼 수 있다.

$$f_{T,U}(x_1, u_2 \mid z) = \lambda_{01}\lambda_{02} e^{x_1\alpha + 2z\beta} e^{-(\lambda_{01}+\lambda_{02})e^{z\beta}x_1 - \lambda_{02}e^{x_1\alpha + z\beta(u_2 - x_1)}} I(u_2 > x_1)$$

$$= f_T(x_1 \mid z) f_{U\mid T}(u_2 \mid x_1, z).$$

단, $f_T(x_1 \mid z) = P(T = x_1, U > x_1)$

$$= \pi\lambda e^{\lambda x_1}.$$

$$\pi = \frac{\lambda_{01}}{\lambda_{01} + \lambda_{02}},$$
$$\lambda = (\lambda_{01} + \lambda_{02})\, e^{z\beta},$$
$$f_{U|T}(u_2 \mid x_1, z) = \lambda^* e^{-\lambda^*(u_2 - x_1)} I(u_2 > x_1),$$
$$\lambda^* = \lambda_{02}\, e^{x_1\alpha + z\beta}.$$

U의 밀도함수는 다음과 같다.

$$f_U(u_1 \mid z) = (1-\pi)\,\lambda\, e^{-\lambda u_1}.$$

모의실험에 필요한 표본들은 다음과 같이 발생하여 사용하였다.

[1] Z의 값 z는 $p = 0.5$인 베르누이 분포로부터 얻었다.

[2] u의 값은 일양분포 (0, 1)로 부터 얻었다.

[3] (a) $u < \pi$이면 x_1값은 모수가 λ인 지수분포로부터 얻고

u_2값은 $f_{U|T}(u_2 \mid x_1, z)$로 부터 얻었다.

(b) $u > \pi$이면 u_1값은 모수가 λ인 지수분포로부터 얻었다.

[4] C의 c값은 모수가 λ_c인 지수분포로부터 얻었다.

$l(l = 1, \cdots, N)$번째 반복 실험에서 $\boldsymbol{\theta} = (\beta, \alpha)$의 최대 편우도 추정량은 뉴톤-랩슨(Newton-Raphson)법을 사용하여 편우도 함수를 최대화하는 것으로 얻어진다.

4.2 모의실험의 결과 및 결론

<표 1>에서 제시한 것처럼 발생 가능한 모든 경우를 4개의 범주로 나누었다. $P_i(z)$를 i번째 범주에 대한 $Z = z$가 주어진 조건부확률이라고 하자. 즉,

$$P_i(z) = P(\text{범주 } i\text{에서 나온 결과} \mid z), \quad i = 1, \cdots, 4.$$

i번째 범주에 대한 비조건부확률은 다음과 같다.

$$P_i = \sum_z p(z)\, P_i(z), \quad i = 1, \cdots, 4.$$

단, $p(z)$는 Z에 대한 확률질량함수이다.

각 범주에 속할 확률은 다음과 같다.

$$P_1(z) = \frac{\lambda_c}{\lambda + \lambda_c}, \quad P_2(z) = (1-\pi)\frac{\lambda}{\lambda + \lambda_c},$$
$$P_{3(z)} = \frac{\pi\lambda}{\lambda + \lambda_c}\ \frac{\lambda_{02}\exp(z\beta)}{\lambda_c(\lambda_c + \lambda - \alpha)}.$$

$$P_4(z) = \frac{\pi\lambda}{\lambda + \lambda_c} \ [1 - \frac{\lambda_{02}\exp(z\beta)}{\lambda_c(\lambda_c + \lambda - \alpha)}].$$

정보적 중도절단 자료를 임의 중도절단 자료로 고려할 때 β의 최대 편우도 추정량의 변화를 조사하기 위하여 모의실험을 하였다. α는 0.7이고 β는 -0.7, -0.3, 0, 0.3, 0.7인 경우 표본 크기를 50과 100으로 하여 각각 1000번씩 반복 실험을 하였다.

정보적 중도절단 자료를 임의 중도절단 자료로 고려한 경우에 편의의 감소는 아래와 같이 계산하였다.

$$Red\_Bias = \left(\frac{Bias_N - Bias_I}{Bias_N}\right) \times 100 ,$$

단, $Bias_N$은 정보적 중도절단 자료를 고려하지 않은 경우의 편의이고, $Bias_I$는 정보적 중도절단 자료를 고려한 경우의 편의이다.

정보적 중도절단 자료를 고려한 경우에 평균제곱오차의 감소는 아래와 같이 계산하였다.

$$Red\_MSE = \left(\frac{MSE_N - MSE_I}{MSE_N}\right) \times 100 ,$$

단, MSE_N은 정보적 중도절단 자료를 고려하지 않은 경우의 평균제곱오차이고, MSE_I는 정보적 중도절단 자료를 고려한 경우의 평균제곱오차이다.

정보적 중도절단 자료를 고려한 경우에 95% 신뢰구간의 평균 폭에 대한 감소는 아래와 같이 계산하였다.

$$Red\_Mean\ W = \left(\frac{Mean\ W_N - Mean\ W_I}{Mean\ W_N}\right) \times 100 ,$$

단, $Mean\ W_N$은 정보적 중도절단 자료를 고려하지 않은 경우의 95% 신뢰구간의 평균 폭이고, $Mean\ W_I$는 정보적 중도절단 자료를 고려한 경우의 95% 신뢰구간의 평균 폭이다.

일반적으로 소아 골수암 연구에서 환자의 대부분이 골수이식과 재발 또는 사망 순으로 경험하기 때문에 P_4가 가장 큰 경우를 고려하였고 좀 더 타당성 있는 결론을 유도하기 위하여 환자의 비율이 네 가지 경우 비슷한 경우를 고려하였다.

<표2>와 <표4>는 P_4가 가장 큰 경우 정보적 중도절단 자료를 임의 중도절단 자료로 잘못 다루었을 때 α와 β의 편의와 평균제곱오차 그리고 95% 신뢰구간의 평균 폭의 감소를 나타낸 것이다. β의 평균제곱오차 감소와 신뢰구간의 평균 폭의 감소는 각각 평균적으로 약 28.4%와 약 15%가 감소함을 알 수 있다. Red_Bias는 크지만 편의는 대체적으로 아주 작으므로 편의의 감소 또는 증가의 폭이 크더라도 별 의미가 없다고 할 수 있다. 정보적 중도절단 자료를 고려한 경우에도 전체적으로 편의, 평균제곱오차, β의 신뢰구간의 평균 폭이 작아지므로 β의 최대 편우도 추정량 방법의 효율은 크다고 볼 수 있다. α의 평균제곱오차의 감소와 95% 신뢰구간의 평균 폭의 감소는 β에 관계없이 각각 약 2.5% 미만과 약 0.6%

미만으로 감소한다.

<표3>과 <표5>는 정보적 중도절단 자료를 고려한 경우와 그렇지 않고 임의 중도절단 자료로 다룬 경우 α와 β의 95% 신뢰구간에 포함되는 확률을 나타낸 것으로 두 경우 모두 비교적 95%에 근접함을 알 수 있다.

반면에 α의 최대 편우도 추정량 방법의 효과는 그다지 크다고 할 수 없다.

<표6>과 <표8>은 $P_1 = P_2 = P_3 = P_4$인 경우에 α와 β의 편의와 평균제곱오차 그리고 95% 신뢰구간의 평균 폭의 감소를 나타낸 것이다. 평균제곱오차의 감소와 신뢰구간의 평균 폭의 감소는 평균적으로 각각 약 55%와 약 32%가 감소함을 알 수 있다. 정보적 중도절단 자료를 고려한 경우에 전체적으로 편의, 평균제곱오차, 신뢰구간의 평균 폭이 아주 많이 감소하므로 β의 최대 편우도 추정량 방법의 효과는 크다고 볼 수 있다. α의 평균제곱오차의 감소와 95% 신뢰구간의 평균 폭의 감소는 평균적으로 각각 약 2.3%와 약 1.2%가 감소한다.

<표7>과 <표9>는 정보적 중도절단 자료를 고려한 경우와 그렇지 않고 임의 중도절단 자료로 다룬 경우 α와 β의 95% 신뢰구간에 포함되는 확률을 나타낸 것으로 n=100이고 $\beta = 0.3$인 경우만 제외하고는 비교적 95%에 근접함을 알 수 있다.

모의실험을 통하여 정보적 중도절단을 바르게 분류한 경우에는 정보적 중도절단 자료를 임의적 중도절단 자료로 잘 못 분류한 경우보다 좀 더 효율적인 α의 신뢰구간에는 큰 변화가 없었지만 평균적으로 β의 신뢰구간은 폭이 평균적으로 30% 내외로 감소됨을 알 수 있었다. 일반적으로 서로 다른 집단간의 위험도는 회귀계수의 추정치에 영향을 받으므로 정보적 중도절단을 바르게 분류하는 것이 효율적으로 위험도를 추정하는데 매우 중요하다고 할 수 있다

<표2> P_4가 가장 큰 경우, 감소된 β의 편의, 평균제곱오차, 95% 신뢰구간 폭

β	n	편 의		평균제곱오차		신뢰구간 평균 폭	
		$Bias_I$	Red_Bias	MSE_I	Red_MSE	$Mean\,W_I$	$Red\_Mean\,W$
-0.7	50	-0.0258	28.7024	0.0949	25.0807	1.1515	16.5754
	100	-0.0187	5.4831	0.0382	29.4647	0.7783	15.3704
-0.3	50	-0.0156	22.4537	0.0825	31.2625	1.0742	15.9922
	100	-0.0051	20.7121	0.0326	29.0721	0.7288	14.6351
0	50	0.0134	-62.736	0.0746	25.4295	1.046	15.5349
	100	0.0071	-722.9195	0.0331	28.8245	0.7107	14.4296
0.3	50	0.0055	47.0401	0.0731	26.7987	1.0414	15.3771
	100	0.0086	-0.6039	0.0364	29.2426	0.7103	14.3955
0.7	50	0.0179	30.0547	0.0847	33.6879	1.0771	15.8899
	100	0.0178	37.969	0.0378	25.3195	0.7317	14.3903

<표3> P_4가 가장 큰 경우, β의 95% 신뢰구간에 포함되는 확률

β	n	정보적 중도절단으로 고려한 경우	임의적 중도절단으로 고려한 경우
-0.7	50	0.951	0.958
	100	0.944	0.956
-0.3	50	0.938	0.948
	100	0.964	0.955
0	50	0.95	0.946
	100	0.951	0.95
0.3	50	0.949	0.959
	100	0.94	0.942
0.7	50	0.947	0.947
	100	0.951	0.949

<표4> P_4가 가장 큰 경우, 감소된 α의 편의, 평균제곱오차, 95% 신뢰구간 평균 폭

β	n	편 의		평균제곱오차		신뢰구간 평균 폭	
		$Bias_I$	Red_Bias	MSE_I	Red_MSE	$Mean\ W_I$	$Red\_Mean\ W$
-0.7	50	0.0566	0.6522	0.0813	2.4772	0.9699	0.586
	100	0.034	0.419	0.0248	0.806	0.5653	0.2879
-0.3	50	0.0478	-2.6155	0.0852	1.4546	0.964	0.369
	100	0.0141	-3.9374	0.0233	0.3697	0.5595	0.1865
0	50	0.0597	-1.7697	0.0875	0.8314	0.9903	0.2995
	100	0.0197	-1.3234	0.025	0.4269	0.5764	0.1288
0.3	50	0.0692	-1.3968	0.0947	1.2112	1.0203	0.3552
	100	0.0244	-3.5561	0.0259	0.8144	0.5948	0.1607
0.7	50	0.0551	1.6545	0.0968	1.925	1.0673	0.5937
	100	0.0155	2.842	0.0264	1.6558	0.6217	0.3436

<표5> P_4가 가장 큰 경우, α의 95% 신뢰구간에 포함되는 확률

β	n	정보적 중도절단으로 고려한 경우	임의적 중도절단으로 고려한 경우
-0.7	50	0.947	0.949
	100	0.955	0.95
-0.3	50	0.963	0.957
	100	0.937	0.936
0	50	0.945	0.943
	100	0.951	0.949
0.3	50	0.944	0.942
	100	0.965	0.965
0.7	50	0.949	0.952
	100	0.947	0.944

<표6> $P_1 = P_2 = P_3 = P_4$가 가장 큰 경우, 감소된 β의 편의, 평균제곱오차, 95% 신뢰구간 평균 폭

β	n	편 의		평균제곱오차		신뢰구간 평균 폭	
		$Bias_I$	Red_Bias	MSE_I	Red_MSE	$Mean\ W_I$	$Red\_Mean\ W$
-0.7	50	-0.0094	80.3223	0.2517	48.0858	2.0018	30.872
	100	-0.0268	64.1188	0.1357	57.8238	1.408	32.7803
-0.3	50	-0.0099	77.2517	0.2551	48.4032	1.8466	30.9528
	100	0.0032	69.8814	0.1096	56.7901	1.2731	31.7394
0	50	-0.0067	53.5499	0.2061	59.6846	1.7478	32.2663
	100	0.0059	-152.3605	0.0911	55.4234	1.2047	31.2401
0.3	50	-0.0073	-283.5673	0.1961	54.051	1.6791	31.6821
	100	-0.0218	26.2242	0.0828	57.3956	1.1261	30.7806
0.7	50	0.0155	69.3316	0.1763	56.2375	1.6339	32.0626
	100	0.0028	84.9763	0.0819	54.7758	1.1142	30.383

<표7> $P_1 = P_2 = P_3 = P_4$인 경우, β의 95% 신뢰구간에 포함되는 확률

β	n	정보적 중도절단으로 고려한 경우	임의적 중도절단으로 고려한 경우
-0.7	50	0.965	0.976
	100	0.962	0.964
-0.3	50	0.945	0.976
	100	0.946	0.954
0	50	0.959	0.963
	100	0.961	0.968
0.3	50	0.955	0.973
	100	0.939	0.934
0.7	50	0.959	0.971
	100	0.952	0.954

<표8> $P_1 = P_2 = P_3 = P_4$가 가장 큰 경우, 감소된 α의 편의, 평균제곱오차, 95% 신뢰구간 평균 폭

β	n	편 의		평균제곱오차		신뢰구간 평균 폭	
		$Bias_I$	Red_Bias	MSE_I	Red_MSE	$Mean\ W_I$	$Red\_Mean\ W$
-0.7	50	-0.282	3.1599	1.8494	1.9583	8.1972	1.6737
	100	-0.0956	-1.9754	1.2876	2.2734	5.2067	1.0532
-0.3	50	-0.141	5.6886	1.6903	4.231	8.15	1.8342
	100	-0.0195	3.7953	1.1715	1.3032	4.6683	0.8709
0	50	-0.2815	5.4186	1.7483	3.9669	7.3064	2.1217
	100	0.0342	-6.3244	1.0153	1.4126	4.3005	0.8846
0.3	50	-0.2093	1.3168	1.7092	2.1678	7.1621	1.1503
	100	-0.0616	-4.445	0.9556	1.5731	3.9502	0.5789
0.7	50	-0.2116	3.457	1.767	2.8197	7.0808	1.3038
	100	-0.0121	8.3966	0.8526	1.2983	3.72	0.629

<표9> $P_1 = P_2 = P_3 = P_4$인 경우, α의 95% 신뢰구간에 포함되는 확률

β	n	정보적 중도절단으로 고려한 경우	임의적 중도절단으로 고려한 경우
-0.7	50	0.997	0.995
	100	0.988	0.989
-0.3	50	0.996	0.996
	100	0.985	0.986
0	50	0.996	0.995
	100	0.982	0.98
0.3	50	0.994	0.991
	100	0.957	0.959
0.7	50	0.995	0.994
	100	0.978	0.978

참고문헌

[1] Fleming, T. R. & Harrington, D. P.(1991). *Counting Processes and Survival Analysis,* New York, Wiley .

[2] Kalbfleisch, J. D. & Prentice, R. L.(1980). *The Statistical Analysis of Failure Time Data*, New York, Wiley.

[3] Klein, J. P., Keiding,N., & Kamby, C., (1989). Semiparametric Marshall-Olkin Models Applied to the Occurrence of Metastases at Multiple Sites after Breast Cancer, *Biometrics*, Vol.45, 1073-1086.

[4] Lagakos, S. W., (1979). General Right Censoring and Its Impact on the Analysis of Survival Data, *Biometrics*, Vol.35. 139-156.

[5] Marshall, A. W., & Olkin, I., (1967). A Multivariate Exponential Distribution, *Journal of the American Statistical Association*, Vol 62, .30-44.

[6] Nádas, A., (1971). The Distribution of the Identified Minimum of a Normal Pair Determines the Distribution of the Pair, *Technometrics*, Vol.13 .201-202.

[7] Smith, F. B. & Helms, R. W.,(1995). EM Mixed Model Analysis of Data from Informatively Censored Normal Distributions, *Biometrics*, Vol. 51. 425-436.