

## Fitting competing risks models using medical big data from tuberculosis patients

Gyeong Dae Kim<sup>a</sup> · Maeng Seok Noh<sup>a</sup> · Chang Hoon Kim<sup>b</sup> · Il Do Ha<sup>a,1</sup>

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

(Received June 28, 2018; Revised July 9, 2018; Accepted July 9, 2018)

---

### Abstract

Tuberculosis causes high morbidity and mortality. However, Korea still has the highest tuberculosis (TB) incidence and mortality among OECD countries despite decreasing incidence and mortality due to the development of modern medicine. Korea has now implemented various policy projects to prevent and control tuberculosis. This study analyzes the effects of public-private mix (PPM) tuberculosis control program on treatment outcomes and identifies the factors that affecting the success of TB treatment. We analyzed 130,000 new tuberculosis patient cohort from 2012 to 2015 using data of tuberculosis patient reports managed by the Disease Control Headquarters. A cumulative incidence function (CIF) compared the cumulative treatment success rates for each factor. We compared the results of the analysis using two popular types of competition risk models (cause-specific Cox's proportional hazards model and subdistribution hazard model) that account for the main event of interest (treatment success) and competing events (death).

Keywords: tuberculosis, public-private mix (PPM), competing event, cumulative incidence function, PPM tuberculosis control program

---

### 1. 서론

결핵은 높은 이환과 사망을 일으키는 질병으로 2017년 세계 보건 기구(World Health Organization)의 보고에 따르면 2016년에는 전 세계적으로 새롭게 결핵으로 보고된 환자가 약 630만 명, 사망자가 167만 명으로 나타났다. 전 세계적으로 결핵 발생률은 연간 약 2%씩 감소하고 있으며 사망률 역시 연간 약 3%씩 감소하는 추세이다 (World Health Organization, 2017). 한국 역시 결핵 발생률과 사망률이 감소하는 추세이지만 아직까지 OECD 국가 중 가장 높은 발생률과 사망률을 보유하고 있다 (Korea Center for Disease Control and Prevention, 2017). 이에 따라 한국은 결핵 관리를 위해 가족접촉자검진사업, 집단시설 역학조사, 결핵환자 진료비지원사업, 입원격리 명령환자 지원사업, 민간공공협력(public-private mix; PPM) 결핵관리사업 등을 실시하고 있다.

여러 결핵 사업 중 PPM 결핵관리사업은 결핵 환자의 관리 및 치료 결과 향상을 위한 정책으로 주요 내용과 목적은 다음과 같다. 먼저 정부에서 추진하고 있는 결핵 사업에 대해 교육을 받고 결핵에 대한

---

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2017R1E1A1A03070747).

<sup>1</sup>Corresponding author: Department of Statistics, Pukyong National University, 45 Yongso-ro, Nam-gu, Busan 48513, Korea. E-mail: [idha1204@gmail.com](mailto:idha1204@gmail.com)

전문지식을 가지고 있는 결핵관리전담간호사를 민간 의료기관에 배치한다. 결핵관리전담간호사는 환자의 복약 관리 및 부작용 상담 등을 통해 환자를 관리한다. 이를 통해 장기간 진행되는 결핵치료에 대해 환자들의 치료 순응도를 높여 치료결과를 향상시키는 것이 최종 목적이다. 한국의 결핵환자는 이전에는 보건소를 많이 이용하였으나 국민 건강보험이 도입된 이후 민간 의료기관의 이용률이 증가하고 있다. 이에 따라 PPM 결핵관리사업이 결핵치료에 대한 한국의 핵심 사업으로 추진되고 있다. 그러나 한국의 PPM에 대한 연구는 아직까지 많지 않다.

이에 따라 본 논문은 PPM 결핵관리사업이 결핵 치료에 미치는 영향을 연구하였으며 결핵환자의 치료 성공과 관련된 요인을 탐색하였다. 본 논문과 기존 연구와의 관련성과 차별성을 설명하기 위해 국내외의 연구 동향을 소개한다. Okanurak 등 (2008)은 방콕의 코호트 자료를 사용하여 로지스틱 회귀 분석을 수행하였으며 이를 통해 결핵 치료 성공에 기여하는 요인을 확인하였다. Falzon 등 (2005)는 유럽 연합 국가의 결핵 환자를 대상으로 로지스틱 회귀 분석을 통해 국가별 치료 성공의 차이를 비교하였으며 치료 성공의 결정 요인을 확인하였다. Park (2011)은 한국의 PPM기관에서 치료 받은 환자와 대조군 환자의 치료결과를 카이제곱 검정을 실시하여 비교하였다. 그러나 기존 연구들은 치료 결과 분석에 있어 성공과 중단을 분리하여 분석을 진행하였으며 생존시간 즉, 치료기간을 고려하지 않았다.

본 논문의 목적은 치료기간, 추적소실 등을 함께 고려하여 PPM의 효과와 치료 성공과 관련된 요인을 탐색하는 것에 있다. 이를 위해 치료기간과 추적소실을 동시에 고려할 수 있는 생존분석 방법을 적용하였으며 경쟁사건(competing event)을 고려하기 위해 경쟁위험모형(competing risks model)을 사용하여 한국의 실제 결핵 데이터를 통해 결과를 확인한다.

본 논문은 다음과 같이 구성되어 있다. 제 2절에서는 분석에 이용한 자료와 변수에 대해 설명한다. 제 3절에서는 본 논문에서 사용한 분석방법에 대해 간략히 소개한다. 제 4절에서는 실제 결핵 데이터에 대한 분석 결과를 확인한다. 제 5절에서는 결론 및 향후 연구에 대해서 논의한다. 제 6절에서는 분석에 사용한 코드를 제시한다.

## 2. 데이터

### 2.1. 자료원

본 논문에서는 질병관리본부로부터 제공받은 한국의 전국 결핵환자 신고 자료를 사용하였다. Son (2018)의 정의에 따라 2012-2015년 전국 결핵 신환자 코호트 137,901명을 확인하였으며 이 중 치료기간 또는 연령을 알 수 없는 4,662명을 제외하고 최종 133,239명의 전국 결핵 신환자를 대상으로 분석을 수행하였다. 본 논문에서 사용한 자료는 질병관리본부와의 정책용역사업을 수행함에 따라 제공받았다.

### 2.2. 자료의 구조 및 설명

치료 결과가 완료, 완치이면 치료성공으로 정의하였으며 치료성공을 관심사건(Type 1)으로 설정하였다. 경쟁사건(Type 2)은 사망으로 설정하였고 중도절단은 결측, 기타, 실패, 전원, 중단, 치료 중으로 설정하였다. 분석에 사용한 설명변수는 연령, 성별, 결핵 유형(폐외결핵, 도말양성 및 폐결핵, 기타 폐결핵) 등의 환자의 인적특성을 고려하였다. 환자가 거주하는 지역의 특성을 고려하기 위해 지역박탈지수를 사용하였다. 지역박탈지수는 지역의 사회경제적 수준을 반영하는 지표로서 건강불평등을 설명하기 위해 사용된다 (Yoon, 2010). 박탈지수가 높을수록 주거환경은 좋지 않다. PPM 효과를 확인하기 위해 치료결과를 신고한 의료기관의 유형(의원, PPM 상급종합병원, 비 PPM 상급종합병원, PPM 병원, 비 PPM 병원, 보건소)을 설명변수로 사용하였다. 환자의 치료 순응도 영향을 확인하기 위해 결핵환자로 신고 된 횟수를 사용하였으며 시기별 치료 효과를 확인하기 위해 환자가 신고 된 연도를 사용하였다.

**Table 2.1.** Basic statistics of the tuberculosis data

Variable	Basic statistic	
Treatment time	Mean	7.49
	Median	6.53
	Range	0.03–63.07
Status	Success(Type 1)	111,474(83.67)
	Death(Type 2)	9,741(7.31)
	Censoring	12,024(9.02)
Deprivation index	Mean	-0.37
	Median	-0.45
	Range	-2.64–2.08
Age	0–9	247(0.19)
	10–19	5,355(4.02)
	20–44	41,817(31.39)
	45–59	31,370(23.54)
	60–74	28,158(21.13)
	≥ 75	26,292(19.73)
Gender	Male	76,100(57.12)
	Female	57,139(42.88)
Year	2012	35,978(27.00)
	2013	34,166(25.65)
	2014	32,995(24.76)
	2015	30,100(22.59)
Type of tuberculosis	Otherpul	28,046(21.05)
	Extrapul	65,017(48.80)
	Smear+ pul	40,176(30.15)
Number of times notified	1	114,953(86.28)
	2	15,352(11.52)
	≥ 3	2,934(2.20)
Medical institution	Clinic	3,861(2.90)
	GH(N)	16,336(12.26)
	GH(Y)	87,391(65.60)
	H(N)	10,655(8.00)
	H(Y)	2,217(1.66)
	Health	12,779(9.59)

GH(N) = non PPM general hospital; GH(Y) = PPM general hospital; H(N) = non PPM hospital; H(Y) = PPM hospital; PPM = public-private mix.

Table 2.1은 분석자료에 대한 기초 통계량이다. 이산형 자료의 경우 빈도 및 퍼센트를 이용하여 요약하였으며, 연속형 자료의 경우 평균, 중앙값, 범위를 이용해 요약하였다. 특히 관심사건인 치료성공의 경우 83.67%로 나타났으며 경쟁사건인 사망의 경우 7.31%, 중도절단 비율은 9.02%로 나타났다.

### 3. 방법론

경쟁사건은 다양한 분야에서 발생할 수 있다. 예를 들어, 유방암의 재발률을 연구하고자 할 때 재발은 관심사건이 된다. 하지만 재발 외에도 다른 사건이 발생 할 수 있다. 예를 들어, 재발 전에 사망이 발생할 수 있으며 이 경우 사망이 경쟁사건이 된다. 경쟁사건이 발생할 경우 관심사건의 관측을 방해하게

된다. 따라서 이와 같이 관심사건에 대해 분석할 때 경쟁사건이 존재한다면 이를 고려할 필요가 있다. 널리 알려진 생존 분석 방법으로는 콕스 비례위험모형(Cox's proportional hazards model; Cox's PH model)이 있다 (Cox, 1972). 그러나 콕스 비례위험모형은 경쟁사건을 고려하지 않으며 이를 중도절단으로 취급한다. 이에 따라 본 논문에서는 경쟁사건을 고려한 생존분석 방법을 사용하였으며 이 절에서는 각 방법에 대해 설명한다.

### 3.1. 누적 발생 함수

누적 발생 함수(cumulative incidence function; CIF)는 경쟁 사건을 고려하여 누적 발생률을 추정하는 방법이다 (Kalbfleisch와 Prentice, 1980). CIF의 정의는 다음과 같다.

$$F_k(t) \equiv P(T \leq t, \epsilon = k) = \int_0^k f_k(t) dt,$$

여기서  $\epsilon$ 은 사건의 유형( $\epsilon \in \{1, 2, \dots, K\}$ ),  $f_k(t)$ 는  $k$  사건에 대한 부차밀도함수(subdistribution density function)이다. 즉,  $F_k(t)$ 는  $t$ 시점까지  $k$  사건이 발생할 확률을 의미한다.

CIF 추정량은 다음과 같이 표현된다.

$$\hat{F}_k(t) = \sum_{i:t_{(i)} \leq t} \hat{S}(t_{(i)}) \frac{d_{ik}}{R_i},$$

여기서  $t_{(i)}$ 는 생존 시간의 순서 통계량( $i = 1, 2, \dots, D_k$ ),  $R_i$ 는  $t_{(i)}$ 시점에서 위험 대상자 수,  $d_{ik}$ 는  $t_{(i)}$ 시점에서  $k$  사건 발생 횟수( $k = 1, 2, \dots, K$ ),  $\hat{S}(t_{(i)})$ 는  $t_{(i)}$ 시점 이전의 모든 사건에 대한 카플란-마이어 추정량 (Kaplan과 Meier, 1958)이다. 일반적으로 누적 발생 함수를 추정할 때 카플란-마이어(Kaplan-Meier; KM) 방법을 많이 사용한다. KM 방법은 생존 함수를 추정하는 방법으로 1-KM으로 나타내면 누적 발생 함수를 구할 수 있다. KM 방법은 경쟁사건을 중도절단으로 간주하여 추정하는 방법으로 1-KM 추정치가 CIF 추정치 보다 과다 추정하게 된다 (Gooley 등, 1999). 본 논문에서도 결핵환자들의 누적 치료성공률을 추정하였을 때 1-KM 추정치가 CIF 추정치보다 과다 추정되어 나타났다 (Figure 3.1). 이에 따라 본 논문에서는 누적 치료성공률을 추정하기 위해 CIF 방법을 사용하였다

### 3.2. 원인별 콕스 비례위험모형

원인별 콕스 비례위험모형(cause-specific Cox's PH model; CSM)에서 위험함수는 다음과 같이 정의된다 (Prentice 등, 1978).

$$\lambda_k(t) = \lim_{\Delta t \rightarrow 0} \frac{\{P(t \leq T \leq t + \Delta t, \epsilon = k | T \geq t)\}}{\Delta t} = \frac{f_k(t)}{S(t)}. \quad (3.1)$$

이때  $\epsilon$ 은 사건의 유형( $\epsilon \in \{1, 2, \dots, K\}$ ),  $S(t)$ 는  $t$ 시점 이전의 모든 사건에 대한 생존함수이다. CSM 방법으로 사건 1에 대한 위험함수를 정의할 때  $t$ 시점 이전에 사건 2가 발생한 개체는  $t$ 시점에서 사건 1에 대한 위험 대상자에서 제외되며 중도절단으로 간주된다 (Kim, 2016). 식 (3.1)에 근거한 CSM은 다음과 같다.

$$\lambda_k(t; x) = \lambda_{0k}(t) \exp(x^T \beta_k),$$

여기서  $\lambda_{0k}(t)$ 는 사건  $k$ 에 대한 기저 위험함수( $k = 1, 2, \dots, K$ ),  $x = (x_1, \dots, x_p)^T$ 는  $p$ 차원 공변량,  $\beta_k = (\beta_{k1}, \dots, \beta_{kp})^T$ 는 사건  $k$ 에 대한 회귀 계수이다.

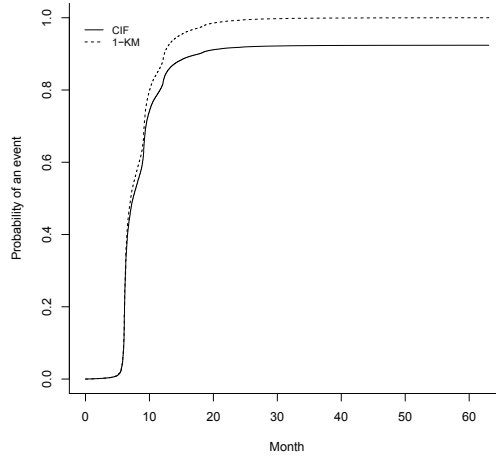


Figure 3.1. Comparison of estimated Kaplan-Meier and cumulative incidence function in the tuberculosis data.

### 3.3. 부차 위험모형

부차 위험모형(subdistribution hazard model; SHM)에서 위험함수는 다음과 같이 정의된다 (Fine과 Gray, 1999).

$$\lambda_k^s(t) = \lim_{\Delta t \rightarrow 0} \frac{\{P(t \leq T \leq t + \Delta t, \epsilon = k | T \geq t \text{ or } (T < t \cap \epsilon \neq k))\}}{\Delta t} = \frac{f_k(t)}{1 - F_k(t)}. \quad (3.2)$$

이때  $F_k(t)$ 는 CIF이다. SHM에서의 위험함수는 CSM의 위험함수와 달리 사건 1에 대한 위험함수를 정의할 때  $t$  시점 이전에 사건 2가 발생한 개체들 역시  $t$  시점에서 사건 1에 대한 위험 대상자에 포함된다 (Kim, 2016). 즉, CSM과 SHM은 서로 다른 위험 대상자 집합(risk set)과 위험 함수를 정의하기 때문에 CSM은 특정 원인을, SHM은 CIF를 각각 모형화 하는 방법이다 (Ha 등, 2017). 식 (3.2)에 근거한 SHM은 다음과 같다.

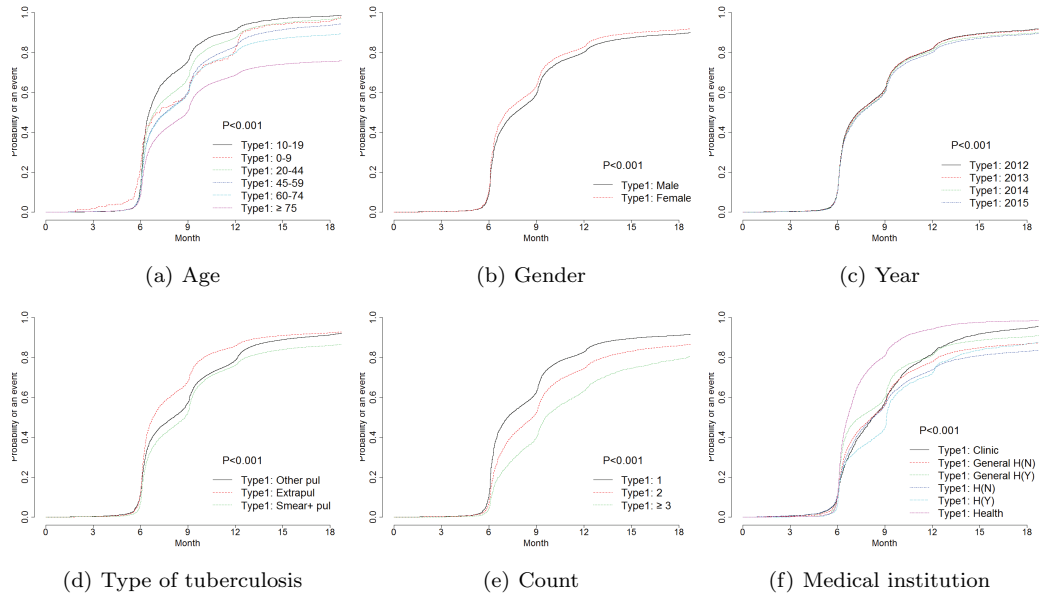
$$\lambda_k^s(t; x) = \lambda_{0k}^s(t) \exp(x^T \beta_k),$$

여기서  $\lambda_{0k}^s(t)$ 는 사건  $k$ 에 대한 기저 부차위험함수이다.

## 4. 분석 결과

Figure 4.1은 각 요인별 누적 치료성공률을 비교한 것이다. 치료기간은 18개월 까지 나타났다. 그레이 검정 (Gray, 1988) 결과 누적 치료성공률은 연령, 성별, 환자가 신고된 연도, 결핵유형, 환자로 신고된 횟수, 치료결과를 신고한 의료기관 유형에 따라 통계적으로 유의하게 차이가 있었다( $p < 0.001$ ).

Table 4.1은 치료성공(Type 1)에 대해 CSM과 SHM을 적합한 결과이다. 위험률(hazard rate)은 특정  $t$ 시점까지는 사건이 발생하지 않았다가  $t$ 시점 바로 직후에 사건이 발생할 순간 확률을 의미한다. 본 논문에서의 관심사건은 치료성공이므로 위험률을 치료성공률로 해석할 수 있다. 따라서 위험비(hazard ratio)가 높을수록 치료성적이 좋게 나타난다. CSM 방법의 경우 박탈지수가 유의하게 나타났으나 SHM 방법에서는 유의하지 않게 나타났다. 또한 CSM 방법은 0-9세를 제외하고는 45-59세까지는 연령이 증가함에 따라 치료성적이 감소하다가 60-74세부터는 다시 치료성적이 증가하여 연령에 따른 뚜렷



**Figure 4.1.** Cumulative treatment success rate for each factor in the tuberculosis data. H(N) = non PPM hospital; H(Y) = PPM hospital; PPM = public-private mix.

한 추세가 나타나지 않는 반면에, SHM 방법에서는 0-9세를 제외하고는 연령이 증가함에 따라 치료성적이 감소하는 경향을 보였다. 두 방법 모두 여자가 남자보다 치료성적이 높았으며 최근 연도에 치료받은 환자일수록 치료성적이 높게 나타났다. 또한 두 방법 모두 폐외결핵 환자의 치료성적이 가장 높았으며 환자로 신고된 횟수가 많을수록 치료성적이 감소하였다. 치료결과를 신고한 의료기관 유형에 따라서는 두 방법 모두 보건소가 가장 높게 나타났다.

## 5. 토의

본 논문에서는 경쟁사건을 고려한 대표적인 모형인 CSM 방법과 SHM 방법을 실제 데이터인 전국 결핵 신환자 자료에 적합하여 결핵관리사업 중 하나인 PPM 결핵관리사업의 효과와 결핵환자의 치료성공에 영향을 주는 요인을 확인하였다. 치료성공에 영향을 주는 요인으로는 SHM 방법에서는 박탈지수가 유의하지 않았으나 CSM 방법에서는 유의하였다. 다른 요인들은 두 방법에서 모두 유의하게 나타났다. 성별, 연도, 환자로 신고된 횟수에 따라서는 두 방법 모두 같은 경향을 나타냈다. 그러나 연령, 결핵유형, 치료결과를 신고한 의료기관유형에 따른 경향은 서로 다르게 나타났다. CSM 방법과 SHM 방법은 서로 다른 위험함수를 설정하기 때문에 서로 다른 모형이므로, 모형선택 등에 대해서는 추후 연구가 필요하다. Lau 등 (2009)에 따르면 CSM 방법은 이미 일어난 사건의 원인을 찾는 데 관심이 있는 경우 적합한 모형 분석법이고, SHM 방법은 예측 측면에서 보다 설명력이 있다고 언급하고 있다.

PPM 결핵관리사업의 효과를 분석하였을 때는 PPM 여부와 상관없이 민간의료기관보다 공공의료기관인 보건소의 치료성적이 가장 높게 나타났다. 이에 따라 민간의료기관에 시행되는 PPM 결핵관리사업의 효과가 좋지 않다고 해석할 수 있으나 본 논문에서는 환자의 동반질환 및 중증도를 고려하지 못하였다. 일반적으로 보건소의 경우 경증의 환자를 관리하여 치료성적이 편이가 생길 수 있으므로 결과해석에 유의하여야 한다. 이에 따라 향후에는 동반질환에 대한 보정 및 중증도를 나타낼 수 있는 공동유무,

**Table 4.1.** Result of CSM and SHM (Type 1)

Variable	CSM		SHM		
	HR	<i>p</i> -value	HR	<i>p</i> -value	
Deprivation index	0.992	0.045	0.997	0.426	
Age (Ref: 10–19)	0–9	0.932	0.318	0.874	0.049
	20–44	0.891	<0.001	0.906	<0.001
	45–59	0.837	<0.001	0.805	<0.001
	60–74	0.877	<0.001	0.710	<0.001
	≥ 75	0.877	<0.001	0.477	<0.001
Gender (Ref: Male)	Female	1.109	<0.001	1.188	<0.001
Year (Ref: 2012)	2013	1.028	0.001	1.020	0.012
	2014	1.070	<0.001	1.030	<0.001
	2015	1.087	<0.001	1.030	<0.001
Type of Tuberculosis (Ref: Other pul)	Extrapul	1.307	<0.001	1.225	<0.001
	Smear+ pul	1.042	<0.001	0.924	<0.001
Number of times notified (Ref: 1)	2	0.711	<0.001	0.773	<0.001
	≥ 3	0.518	<0.001	0.626	<0.001
Medical Institution (Ref: GH(Y))	Clinic	0.829	<0.001	0.929	<0.001
	GH(N)	0.946	<0.001	0.911	<0.001
	H(N)	0.936	<0.001	0.872	<0.001
	H(Y)	0.922	0.001	0.939	0.005
	Health	1.453	<0.001	1.543	<0.001

CSM = cause-specific Cox's proportional hazards model; SHM = subdistribution hazard model; HR = hazard ratio; GH(N) = non PPM general hospital; GH(Y) = PPM general hospital; H(N) = non PPM hospital; H(Y) = PPM hospital; PPM = public-private mix.

객혈여부 등을 고려하여 분석하고자 한다. 여러 제한에도 불구하고 약 13만명의 전국 결핵 신환자 전수 자료를 이용하여 탐색적으로 요인별 효과를 확인한 것에 대해 본 논문의 의의가 있다고 본다.

## 6. 분석 코드

이 절에서는 분석에 사용한 코드(R 및 SAS)에 대해 간략히 설명한다.

### 6.1. 누적 발생 함수

다음은 R을 이용하여 본 논문에서 나타낸 사건 1에 대한 연도별 누적 발생 함수 그림(Figure 4.1(c))을 출력하는 코드이다.

```
library(cmprsk)
fit ← cuminc(time, status, group = year)
inf = list(list(fit$'2012 1'$time, fit$'2012 1'$est),
           list(fit$'2013 1'$time, fit$'2013 1'$est),
           list(fit$'2014 1'$time, fit$'2014 1'$est),
           list(fit$'2015 1'$time, fit$'2015 1'$est))
plot.cuminc(inf, lty=c(1:4))
```

group에는 각 요인을 지정하며 inf에는 요인의 범주 수 만큼 list를 만들어준다. 사건 2에 대한 누적 발생 함수 그림을 출력하기 위해선 1'를 2'로 수정하면 된다.

### 6.2. 원인별 콕스 비례위험모형

다음은 SAS를 이용한 사건 1에 대한 원인별 콕스 비례위험모형 코드이다.

```
proc phreg data = dataname;
class x1 x2 x3 ;
model time*status(0,2) = x1 x2 x3;
run;
```

class 문에는 범주형 변수를 입력한다. 사건 2에 대해 분석하기 위해서는 status(0,2)를 status(0,1)로 수정하면 된다.

### 6.3. 부차 위험모형

다음은 SAS를 이용한 사건 1에 대한 부차 위험모형 코드이다.

```
proc phreg data = dataname;
class x1 x2 x3 ;
model time*status(0) = x1 x2 x3 / eventcode = 1;
run;
```

사건 2에 대해 분석하기 위해서는 eventcode = 1을 eventcode = 2로 수정하면 된다.

## References

- Cox, D. R. (1972). Regression models and life tables (with discussion), *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 187–220.
- Falzon, D., Le Strat, Y., Belghiti, F., Infuso, A., and EuroTB correspondents (2005). Exploring the determinants of treatment success for tuberculosis cases in Europe, *International Journal of Tuberculosis and Lung Disease*, **9**, 1224–1229.
- 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.
- Gooley, T. A., Leisenring, W., Crowley, J., and Storer, B. E. (1999). Estimation of failure probabilities in the presence of competing risks: new representations of old estimators, *Statistics in Medicine*, **18**, 695–706.
- Gray, R. J. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk, *Annals of Statistics*, **16**, 1141–1154.
- Ha, I. D., Jeong, J. H., and Lee, Y. J. (2017). Statistical Modelling of Survival Data with Random Effects: H-likelihood Approach. Springer.
- Kalbfleisch, J. D. and Prentice, R. L. (1980). *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York.
- Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457–481.
- Kim, M. J. (2016). Estimation methods and interpretation of competing risk regression models, *The Korean Journal of Applied Statistics*, **29**, 1231–1246.
- Korea Center for Disease Control and Prevention (2017). Guidelines for National Tuberculosis Control 2017. Cheongju



- Lau, B., Cole, S. R., and Gange, S. J. (2009). Competing risk regression models for epidemiologic data, *American Journal of Epidemiology*, **170**, 244–256.
- Okanurak, K., Kitayaporn, D. and Akarasewi, P. (2008). Factors contributing to treatment success among tuberculosis patients: a prospective cohort study in Bangkok, *International Journal of Tuberculosis and Lung Disease*, **12**, 1160–1165.
- Park J. S. (2011). Increasing the Treatment Success Rate of Tuberculosis in a Private, *Tuberculosis and Respiratory Diseases*, **70**, 143–149.
- 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.
- Son, H.J. (2018). Current Status and Determinants of Treatment Outcomes among New Tuberculosis Patients in Republic of Korea: Achievement in Public-private mix program and challenges, *Hanyang University doctoral dissertation*.
- World Health Organization (2017). Global Tuberculosis report 2017. Geneva.
- Yoon, T. H. (2010). Regional Health Inequalities in Korea The Status and Policy Tasks, *Critical Social Welfare Academy*, **30**, 49–77.

# 전국 결핵 신환자 의료빅데이터를 이용한 경쟁위험모형 적합

김경대<sup>a</sup> · 노맹석<sup>a</sup> · 김창훈<sup>b</sup> · 하일도<sup>a,1</sup>

<sup>a</sup>부경대학교 통계학과, <sup>b</sup>부산대학교병원

(2018년 6월 28일 접수, 2018년 7월 9일 수정, 2018년 7월 9일 채택)

---

## 요약

결핵은 높은 이환과 사망을 일으키는 질병으로 현대의학의 발달에 따라 발생률과 사망률은 감소하고 있다. 그러나 한 국은 아직까지 OECD 국가 중 결핵 발생률과 사망률이 가장 높다. 이에 따라 한국은 결핵의 예방 및 통제를 위해 여러 정책 사업을 실시하고 있다. 본 연구에서는 공공민간협력(public-private mix) 결핵관리사업이 치료결과에 미치는 영향을 분석하고 결핵환자의 치료 성공에 영향을 미치는 요인을 확인하고자 한다. 질병관리본부에서 관리하는 결핵환자 신고 자료를 이용하여 2012-2015년 전국 결핵 신환자 코호트 약 13만명을 대상으로 분석하였다. 누적 발생 함수(cumulative incidence function)를 이용하여 요인별로 누적 치료 성공률을 비교하였으며, 주 관심사건(치료성공) 및 경쟁사건(사망)을 고려한 두 가지 경쟁위험모형(cause-specific Cox's proportional hazards model and subdistribution hazard model)을 사용하여 분석 결과를 비교하였다.

주요용어: 결핵, 공공민간협력, 경쟁사건, 누적발생함수

---

---

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. NRF-2017R1E1A1A03070747).

<sup>1</sup>교신저자: (48513) 부산광역시 남구 용소로 45, 부경대학교 통계학과. E-mail: idha1204@gmail.com